SCHAUM'S ouTlines

全美经典 学习指导系列

统计学原理

（上册）

——描述性统计学与概率

〔美〕S. 伯恩斯坦 R. 伯恩斯坦 著

史道济 译

- 获取高分的最佳助手
- 有关数学知识的总复习
- 469道完全解答的习题
- 涵盖本课程的所有基础，是任一教材的补充
- 理想的自学读物

科学出版社

麦格劳-希尔教育出版集团

(O-1540.0101)

责任编辑：刘嘉善

SCHAUM'S ouTlines

全球销量
超越3000万的
“全美经典学习指导系列”
是您的最佳
学习伴侣！

40年来最畅销的教辅系列

全美著名高校资深教授倾力之作

国内重点高校任课教师全力推荐并担当翻译

省时高效的学习辅导，全面详细的习题解答

迄今为止国内最全面的教辅系列

覆盖大学理工科专业

全美经典 学习指导系列

概率和统计
统计学
离散数学
Mathematica使用指南
数理金融引论
机械振动
微分方程
统计学原理（上）
统计学原理（下）
微积分
静力学与材料力学
有限元分析
传热学
近代物理学

2000工程力学习题精解
工程力学
3000物理习题精解
流体动力学
物理学基础
材料力学
2000离散数学习题精解
工程热力学
数值分析
量子力学
有机化学习题精解
3000化学习题精解
大学化学习题精解
电路

电气工程基础
工程电磁场基础
数字信号处理
数字系统导论
数字原理
电机与机电学
基本电路分析
信号与系统
微生物学
生物化学
生物学
分子和细胞生物学
人体解剖与生理学

[illegible] [illegible]

ISBN 7-03-009771-8

9 787030 097712 >

McGraw Hill

ISBN 7-03-009771-8/O · 1540

定 价：24.00 元

全美经典学习指导系列

统计学原理

（上册）

——描述性统计学与概率

〔美〕S. 伯恩斯坦
R. 伯恩斯坦 著

史道济 译

科学出版社

麦格劳-希尔教育出版集团

2002

内 容 简 介

本书论述工程技术、自然科学和生命科学中常用的统计学原理和方法. 全书分为两册，上册内容包括学习本书所需的数学知识、描述性统计学的基本原则和方法以及推断性统计学的理论基础，概率论.

本书每一章都有相同的形式：第一部分以大纲的形式论述所有的新概念和新方法以及有完整解答的例子. 第二部分是习题解答，包括许多理论的应用. 第三部分是补充习题，只有答案. 这部分内容是检验读者对本书内容的理解程度.

本书内容丰富，论述严谨. 为了自学的方便，全书有相互参照的系统可以很快找到要学习的内容.

本书适合于高等院校理工科教师与学生和有关工程技术人员阅读.

Schaum's Outlines
Stephen Bernstein and Ruth Bernstein: Elements of Statistics I: Descriptive Statistics and Probability
ISBN: 0-07-005023-6

图字：图字 01－2001－2119 号

图书在版编目(CIP)数据

统计学原理(上册)——描述性统计学与概率/〔美〕S. 伯恩斯坦，〔美〕R. 伯恩斯坦著；史道济译. —北京：科学出版社，2002. 1
(全美经典学习指导系列)
ISBN 7－03－009771－8
Ⅰ. 统… Ⅱ. ①伯…②伯…③史… Ⅲ. 统计学 Ⅳ. C8
中国版本图书馆 CIP 数据核字(2001)第 063367 号

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2002 年 1 月第 一 版　　开本：A4(890×1240)
2002 年 1 月第一次印刷　　印张：15 3/4
印数：1—5 000　　字数：447 000

定价：24.00 元

(如有印装质量问题，我社负责调换〈环伟〉)

前　　言

统计学是论述收集、分析并解释数字信息的科学. 对这门科学有一个基本的了解，不仅对每个做研究工作的科学家是重要的，而且对现代社会中许多人都是重要的，他们必须处理这种信息：如评价有争议的医学研究报告的医生，企图使陪审团相信定量证据是合法的律师、改进质量控制程序的制造商，解释市场趋势的经济学家等等.

统计科学的理论基础是数学中称为数理统计的学科，这里统计学是抽象的公理、定理和严谨的证明紧密结合而形成的完整的结构. 为了使这种理论结构也适用于非数学家，产生了一个称为一般统计学的解释性的学科，其中的描述是最简化的，且常常是非数学的. 每个特殊领域（例如农学、人类学、生物学、经济学、工程学、心理学、社会学）都从这个简化的形式得到了适合于各自数据的材料. 例如有一种称为生物统计学的，就是特别适合于生物学数据资料的一般统计学.

在一般统计学或它的某些特殊分支中，所有的介绍性课程都有相同的核心：统计学基础，多年来，本书作者学习了这些课程的基础，在研究项目中也应用过，并且教过一般统计学和生物统计学课程. 在此经历上写成的本书是一本统计学基础的自学指导，可以自己阅读，也可作为教材的补充，而且由于本书内容完整，实际上还可作为教材使用.

统计科学可以分为两部分：描述性统计与推断性统计. 描述性统计提供了将原始数据整理成有用形式的方法，这些方法包括收集、整理、概括、描述及给出数据信息. 如果在研究中始终利用全部资料（总体），那么要求描述性统计. 但有时可以利用的只是总体的一小部分（样本），此时要求由有限的、不确定的样本信息，对整个总体做出判断及决策的方法，这是推断性统计的领域.

在介绍一般统计学的所有课程中，都以标准的次序给出统计学的这两个部分，本书也依这个次序，但将它们分为二册. 上册（第一—第十章）讲述为学习本书所必须的数学（大学代数的概念），描述性统计的基本原则、方法及推断性统计的主要理论基础：概率论. 下册（第十一—第二十章）论述推断性统计的概念和方法. 全书每章都有相同的形式：第一部分是正文，以大纲的形式论述所有的新概念和新方法以及有完整解答的例子. 下一部分是习题解答，复习同样的材料且让你考虑来自不同视角的材料，最后一部分是补充习题，检查你对内容的掌握，只有答案，而没有解答步骤. 因为这是一本关于一般统计的书，我们努力对代表许多特殊领域的问题作出形式多样的选择，而且还希望用这些问题去说明，如何由实际问题——正在解决的事情的数据信息作出决策.

为了掌握统计学，你必须阅读课文并做习题，我们建议你首先阅读正文及随后的例子，然后在做习题解答及补充习题前，回过头来再看一次正文. 这本书全书有相互参照系统，使你可以很快的复习对理解后面内容所需要的先前的内容.

如果你继续学习统计学，你可能会使用计算机及许多统计程序包中的一个，本书并不论述如何使用这类计算程序，但会告诉你为理解使用程序所要求的概念，以及同样重要的，解释计算机输出结果. 不要求用计算机做本书中的习题，所有的习题只用电子计算器即可解决.

我们感谢对本书出版作出特殊贡献的 McGraw-Hill 公司的以下人员：Barbara Gilson，Elizabeth Zayatz，John Aliano，Fred Perkins，Arthur Biderman，Mary Loebing Giles 以及 Meaghan McGovern，我还要感谢 E. Kirk 允许翻印初等统计学第二版的表 D. 10，感谢所有允许我们使用他们出版物的个人和组织（在材料出处特别提到的），我们也感谢各章不知名的评论员.

目　录

第一章　统计学的数学基础 …… 1

1.1　什么是统计学 …… 1

1.2　分数运算 …… 1

1.3　带符号数的运算 …… 2

1.4　舍入运算 …… 2

1.5　绝对值 …… 2

1.6　阶乘 …… 3

1.7　开方和根 …… 3

1.8　平方根运算 …… 3

1.9　幂运算 …… 3

1.10　对数运算 …… 4

1.11　代数表达式 …… 4

1.12　方程和公式 …… 5

1.13　变量 …… 5

1.14　单变量方程和二次公式 …… 6

1.15　统计中的变量 …… 6

1.16　可观测变量,假设变量和测量变量 …… 6

1.17　函数和关系 …… 7

1.18　函数记号 …… 7

1.19　统计中的函数 …… 8

1.20　实数轴和直角笛卡儿坐标系 …… 8

1.21　函数的图象 …… 9

1.22　序列、级数和求和符号 …… 10

1.23　不等式 …… 10

第二章　统计资料的特征 …… 23

2.1　测量尺度 …… 23

2.2　测量的操作定义 …… 23

2.3　测量水平和测量单位 …… 23

2.4　名义水平测量 …… 23

2.5　次序水平测量 …… 24

2.6　间隔水平测量 …… 24

2.7　比例水平测量 …… 24

2.8　连续测量变量和离散测量变量 …… 25

2.9　统计资料的类型 …… 25

2.10　测量的近似性 …… 26

2.11　有效数字 …… 26

2.12　科学记数法和数量级 …… 27

2.13　测量的系统误差和随机误差 …… 27

2.14　统计中的准确度和精密度 …… 27

2.15　自然科学中的准确度和精密度 …… 28

2.16 单位换算 …… 28
第三章 总体、样本和统计量 …… 34
3.1 自然总体和测量总体 …… 34
3.2 有限总体、无限总体和假设总体 …… 34
3.3 样本 …… 34
3.4 参数与统计量 …… 35
3.5 统计科学 …… 35
3.6 估计问题和假设检验问题 …… 36
3.7 统计假设和研究假说 …… 36
3.8 探索性研究和假设检验研究 …… 37
3.9 探索性试验 …… 37
3.10 对照试验 …… 38
3.11 观察式研究 …… 38
3.12 调查和普查 …… 38
3.13 参数和非参数统计方法 …… 39
3.14 数理统计学和一般统计学 …… 40
3.15 描样设计 …… 40
3.16 抽样的概率:有放回和无放回 …… 40
3.17 随机抽样 …… 41
3.18 简单随机抽样 …… 41
3.19 分层随机抽样 …… 42
3.20 系统随机抽样 …… 42
3.21 整群随机抽样 …… 42
3.22 非随机抽样 …… 43
3.23 随机数表 …… 43
第四章 描述性统计:将统计资料整理成表格形式 …… 50
4.1 阵列和极差 …… 50
4.2 频数分布 …… 50
4.3 相对频数分布和百分数分布 …… 51
4.4 分组频数分布 …… 51
4.5 分组频数分布和分组百分数分布 …… 52
4.6 未分组分布转换成分组分布时应遵循的原则 …… 53
4.7 开端点组分布和不等组距 …… 53
4.8 “小于式”累积分布 …… 55
4.9 “大于等于式”累积分布 …… 55
4.10 分组累积分布 …… 56
第五章 描述性统计:统计资料的图形化 …… 69
5.1 柱状图、线型图和饼状图 …… 69
5.2 条形图 …… 69
5.3 直方图:未分组数据 …… 70
5.4 直方图:分组数据 …… 71
5.5 折线图:未分组数据 …… 71
5.6 折线图:分组数据 …… 72
5.7 频数曲线、相对频数曲线和百分数曲线 …… 72
5.8 象形图 …… 72

5.9 饼状图 …… 73
5.10 茎叶表示法 …… 74
5.11 累积分布曲线图 …… 75
第六章 描述性统计:集中趋势、平均值和位置的度量 …… 96
6.1 集中趋势、平均值和位置的度量 …… 96
6.2 算术平均数 …… 96
6.3 算术平均数的舍入准则 …… 97
6.4 与算术平均数的离差和分布重心 …… 98
6.5 平均值的一种度量——算术平均数 …… 98
6.6 由未分组频数分布计算算术平均数 …… 99
6.7 由分组频数分布计算近似算术平均数 …… 99
6.8 由编码数据计算算术平均数 …… 100
6.9 加权平均 …… 101
6.10 总平均 …… 102
6.11 几何平均 …… 102
6.12 调和平均 …… 102
6.13 中位数和其他分位数 …… 103
6.14 阵列的分位数计算公式 …… 103
6.15 未分组频数分布的分位数计算公式 …… 104
6.16 分组频数分布的分位数计算公式 …… 105
6.17 中列数、四分位数中点和三点均值 …… 106
6.18 众数 …… 107
6.19 分组频数分布的众数计算公式 …… 107
第七章 描述性统计:离散性度量 …… 125
7.1 极差作为一种离散性度量为什么具有有限值 …… 125
7.2 平均偏差 …… 125
7.3 平均偏差的频数分布式 …… 126
7.4 近似平均偏差 …… 127
7.5 总体方差:定义式 …… 128
7.6 总体方差:计算式 …… 129
7.7 样本方差:定义式 …… 129
7.8 样本方差:计算式 …… 130
7.9 总体标准差 …… 130
7.10 样本标准差 …… 131
7.11 离散度量中的舍入原则 …… 131
7.12 由非分组频数分布计算标准差 …… 132
7.13 由分组频数分布计算近似标准差 …… 133
7.14 计算编码数据的方差和标准差 …… 134
7.15 Chebyshev 定理 …… 135
7.16 经验法则 …… 136
7.17 集中趋势和偏离性的图示 …… 136
7.18 变异系数 …… 137
7.19 标准分和标准化变量 …… 138
7.20 四分位极差和四分位差 …… 139
7.21 盒子图与五数概括 …… 139

第八章　概率:古典解释,相对频数解释,集合论解释和主观解释 …… 158
8.1　概率的古典解释 …… 158
8.2　概率的相对频数解释 …… 159
8.3　集合,子集和样本空间 …… 160
8.4　事件 …… 161
8.5　Venn 图 …… 162
8.6　概率的集合论解释 …… 163
8.7　概率的主观解释 …… 166
8.8　机会比率的概念 …… 166
8.9　由机会比率确定概率 …… 167
第九章　计算法则和计数法则 …… 180
9.1　事件组合的概率计算 …… 180
9.2　条件概率 …… 180
9.3　一般乘法法则 …… 182
9.4　独立事件和相关事件 …… 183
9.5　特殊乘法法则 …… 183
9.6　一般加法法则 …… 184
9.7　从一般加法法则中导出特殊加法法则 …… 185
9.8　列联表,联合概率表及边缘概率表 …… 186
9.9　BAYES 定理 …… 188
9.10　树型图 …… 189
9.11　计数法则 …… 190
9.12　计数法则:乘法原理 …… 190
9.13　计数法则:排列 …… 192
9.14　计数法则:组合 …… 192
第十章　随机变量,概率分布和累积分布函数 …… 215
10.1　随机变量 …… 215
10.2　离散型与连续型随机变量 …… 216
10.3　离散型概率分布 …… 216
10.4　连续型概率分布 …… 218
10.5　离散型概率分布和描述性分布的关系 …… 220
10.6　连续型概率分布和描述性分布的关系 …… 221
10.7　离散型随机变量的累积分布函数 …… 222
10.8　连续型随机变量的累积分布函数 …… 224
10.9　离散型随机变量的期望值 …… 225
10.10　连续型随机变量的期望值 …… 226
10.11　离散型随机变量的方差和标准差 …… 226
10.12　离散型随机变量的方差和标准差的计算公式 …… 227
10.13　连续型随机变量的方差和标准差 …… 228
10.14　Chebyshev 定理和经验法则 …… 229
附录 …… 239
表 A.1　随机数表 …… 239
表 A.2　统计分类数据 …… 242

第一章　统计学的数学基础

1.1　什么是统计学

统计学是关于数据资料的收集、整理、分析和推断的一门科学. 它可分为**描述统计学**和**推断统计学**两大类. 描述统计学给出的是将原始数据资料加工成有用图表的方法，这些方法包括数据的收集、整理、概括和描述等. 如果在研究中可以得到整个**总体**，那么描述统计学就足够了. 但是，实际中往往只能得到总体的一小部分(称为**样本**)，这就需要通过这些样本的有限的、不确定的信息来确定有关总体的信息，这就是推断统计学的研究领域.

统计学的理论基础是**数理统计学**. 数理统计学是数学的一个分支，由一系列的公理、定理以及严格证明来组成. 它还涉及到数学的其他一些领域，例如微积分、概率论和高等代数等等. 为了使这些理论也适用于一般研究者，人们将其简化，变得非数学化，由此产生了**一般统计学**. 不同的专业领域(如建筑学、人类学、生物学、经济学等等)与一般统计学结合，就产生了相应的专业统计学. 例如，**生物统计学**即为专门适用于对生物学中的数据资料进行分析的一般统计学.

这本书由两册组成，主要通过许多专业领域的实例和问题来介绍一般统计学. 本书采用的是一般统计学课程的典型提纲：首先介绍描述统计学——数据的收集(第二章和第三章)、整理(第四章)、图形化(第五章)和描述(第六章和第七章)；然后介绍概率论(第八章到第十二章)和抽样理论(第十三章)；最后是推断统计学的估计理论和假设检验理论(第十四章到第二十章).

本书要求读者有大学的数学基础，在第一章复习所需的代数基本知识，假定你有一个电子计算器.

1.2　分数运算

若用 m,n 表示两个数，那么 m 与 n 的乘积可以用如下的**等价符号**表示：$m\times n$、$(m)(n)$ 和 $m\cdot n$. 类似地，m 除以 n 的商可以等价地表示为 $m\div n$，m/n 和 $\frac{m}{n}$.

例 1.1　设 $m=4,n=2$，计算下列式子：$m\times n, m\div n, m/n, (m)(n), m\cdot n, \frac{m}{n}$.

解 ☞

$$4\times 2=(4)(2)=4\cdot 2=8$$

$$4\div 2=4/2=\frac{4}{2}=2$$

分数运算中，分子、分母同乘以或除以相同的数，分数的值不变. 但是分子、分母都加上或减去相同的数，分数的值通常会发生变化.

例 1.2　下列哪些分数与 $\frac{6}{8}$ 是等价的：$\frac{8}{9},\frac{12}{16},\frac{2}{4},\frac{3}{4},\frac{16}{18},\frac{1}{2}$.

解 ☞

$$\frac{6}{8}=\frac{(2)(6)}{(2)(8)}=\frac{12}{16}=\frac{6/2}{8/2}=\frac{3}{4}$$

而

$$\frac{6}{8}\neq\frac{6-4}{8-4}=\frac{2}{4}=\frac{1}{2}\text{和}\frac{6}{8}\neq\frac{6+10}{8+10}=\frac{16}{18}=\frac{8}{9}$$

分数的加减运算必须转化为同分母才可以进行. 分数相乘等于分子分母分别相乘. 分数相除等于除数乘以被除数的倒数.

例 1.3　计算下列各式：(a) $\frac{4}{5}+\frac{5}{6}$，(b) $\frac{5}{7}\times\frac{1}{4}\times\frac{2}{3}$，(c) $\frac{f}{g}\div\frac{m}{n}$.

解 ☞　(a) $\frac{4}{5}+\frac{5}{6}=\frac{4\times 6}{5\times 6}+\frac{5\times 5}{6\times 5}=\frac{24}{30}+\frac{25}{30}=\frac{49}{30}=1\frac{19}{30}$

(b) $\frac{5}{7}\times\frac{1}{4}\times\frac{2}{3}=\frac{5\times1\times2}{7\times4\times3}=\frac{10}{84}=\frac{5}{42}$

(c) $\frac{f}{g}\div\frac{m}{n}=\frac{f}{g}\times\frac{n}{m}=\frac{fn}{gm}$

1.3 带符号数的运算

同号数相加，取原来的符号，并把数值相加. 异号数相加，用数值大的减去数值小的，并取数值大的符号. 带符号数相减等于被减数改变符号后相加.

例 1.4 对下列带符号数进行加减运算：(a) $5+7+9+2$，(b) $5+(-7)+9+(-2)$，(c) $(-5)-(+7)+9-(-2)$.

解 ☞ (a) $5+7+9+2=23$

(b) $5+(-7)+9+(-2)=14-9=5$

(c) $(-5)-(+7)+9-(-2)=-5-7+9+2=-1$

两数相乘或相除，同号得正，异号得负.

例 1.5 对下列带符号数进行乘法和除法运算：(a) $(-4)(-2)$，(b) $(-4)(2)$，(c) $(-4)\div(-2)$，(d) $(-4)/(2)$.

解 ☞ (a) $(-4)(-2)=8$

(b) $(-4)(2)=-8$

(c) $(-4)\div(-2)=2$

(d) $(-4)/(2)=-2$

若一个算式中含有加、减、乘、除运算时，先做乘除法，再做加减法. 如果有括号，就先做括号中的运算.

例 1.6 计算：(a) $2\times10-9$，(b) $[(-2)+(-3)]\div[(-2)-(-3)]$.

解 ☞ (a) $2\times10-9=20-9=11$

(b) $[(-2)+(3)]\div[(-2)+(-3)]=[1]\div[-5]=-0.2$

1.4 舍入运算

用舍入原则进行取整运算中，若小数部分小于 0.5 则舍去小数部分，整数部分保持不变；若小数部分大于 0.5 则去掉小数部分，并且整数部分加 1；若小数部分恰好等于 0.5，则通常采用如下原则：当个位数为奇数时，整数部分加 1，当个位数为偶数时，整数部分保持不变.

例 1.7 对下列数做取整运算：(a) 2.2，(b) 1.89，(c) 2.5，(d) 1.50.

解 ☞ (a) 2.2 舍入到 2

(b) 1.89 舍入到 2

(c) 2.5 舍入到 2

(d) 1.50 舍入到 2

将数舍入到一个小数位置的运算与上述取整运算的基本原则相同，只是用在小数位置以后的小数.

例 1.8 舍入运算：(a) 1.933 舍入二位小数，(b) 0.01791 舍入到二位小数，(c) 1.23915 舍入到三位小数，(d) 0.0015 舍入到三位小数.

解 ☞ (a) 1.933 舍入到 1.93

(b) 0.01791 舍入到 0.02

(c) 1.23915 舍入到 1.239

(d) 0.0015 舍入到 0.002

1.5 绝对值

数 n 的绝对值就是这个数的数值部分，而不考虑符号. 通常用 $|n|$ 表示.

例 1.9 计算下列数的绝对值：(a) -5，(b) $10/2$，(c) $\frac{43-52}{9}$

解 ☞ (a) $|-5|=5$

(b) $|10/2|=5$

(c) $\left|\frac{43-52}{9}\right|=1$

1.6 阶乘

符号 $n!$（读做 n 阶乘）表示从 n 到 1 所有正整数的乘积

$$n! = n\times(n-1)\times(n-2)\times(n-3)\times\cdots\times 1$$

例 1.10 计算下列阶乘：(a) $2!$，(b) $4!$，(c) $9!$.

解 ☞ (a) $2! = 2\times1=2$

(b) $4! = 4\times3\times(2!)=24$

(c) $9! = 9\times8\times7\times6\times5\times(4!)=362,880$

1.7 开方和根

在表达式 $a=\sqrt[n]{b}$ 中，符号 $\sqrt{\ }$ 称为开方符号，$\sqrt[n]{b}$ 称为根式，a 称为 b 的 n 次方根，b 称为被开方数，n 称为根指数.

例 1.11 求下列根：$\sqrt[2]{4}$.

解 ☞ $\sqrt[2]{4}=\sqrt{4}=4$ 的二次根（或平方根）. 4 的平方根是 $+2$，或者 -2. 但是按照惯例我们记 $\sqrt{4}=+2$，$-\sqrt{4}=-2$.

一个数的主 n 次方根是它的一个实数根，或者是正、负根中的正根.

例 1.12 求解主 n 次方根：(a) $\sqrt{16}$，(b) $-\sqrt{16}$.

解 ☞ (a) $\sqrt{16}=4$.

(b) $-\sqrt{16}=-4$.

1.8 平方根运算

两个平方根相乘等于被开方数相乘再开方. 两个平方根相除等于被开方数相除再开方.

例 1.13 计算：(a) $\sqrt{5}\sqrt{5}$，(b) $\frac{\sqrt{125}}{\sqrt{5}}$.

解 ☞ (a) $\sqrt{5}\sqrt{5}=\sqrt{5\times5}=\sqrt{25}=5$

(b) $\frac{\sqrt{125}}{\sqrt{5}}=\sqrt{\frac{125}{5}}=\sqrt{25}=5$

平方根与数相乘等于被开方数乘以此数的平方再开方. 平方根除以一数等于被开方数除以此数的平方再开方.

例 1.14 计算：(a) $2\sqrt{20.25}$，(b) $\frac{\sqrt{25}}{5}$.

解 ☞ (a) $2\sqrt{20.25}=\sqrt{2^2}\sqrt{20.25}=\sqrt{4(20.25)}=\sqrt{81}=9$

(b) $\frac{\sqrt{25}}{5}=\frac{\sqrt{25}}{\sqrt{25}}=\sqrt{\frac{25}{25}}=1$

1.9 幂运算

b^n 是 b 的 n **次幂**，它是 n 个 b 的乘积. 例如，$b^2=b\times b$ 是 b 的二次幂（**b 的平方**），$b^3=b\times b\times b$ 是 b 的三次幂（**b 的立方**）. 在表达式 b^n 中，n 称为**指数**，b 称为**底数**.

例 1.15 计算下列幂运算:(a) b^0, (b) 12^0, (c) 12^2, (d) 5^5.

解 ☞ (a) 任何不等于零的实数的零次幂都等于 1:$b^0=1$

(b) $12^0=1$

(c) $12^2=12\times12=144$

(d) $5^5=5\times5\times5\times5\times5=3,125$

例 1.16 将下列表达式用分数、根式表示:(a) b^{-n}, (b) $b^{1/n}$, (c) $b^{-1/n}$.

解 ☞ (a) 对于任何不等于零的实数,如果它的指数为负数,那么有下列关系成立:$b^{-n}=\frac{1}{b^n}$

(b) $b^{1/n}=\sqrt[n]{b}$

(c) $b^{-1/n}=\frac{1}{\sqrt[n]{b}}$

任何数均可表示为 10 的整数次幂与数 a 的乘积,其中 a 的绝对值大于等于 1 且小于 10,习惯上称之为科学记数法.

例 1.17 用科学记数法表示下列各数:(a) 237, (b) 0.000237, (c) 116,270,000.

解 ☞ (a) $237=2.37\times10^2$

(b) $0.000237=2.37\times10^{-4}$

(c) $116,270,000=1.1627\times10^8$

同底数的幂相乘,指数相加;同底数的幂相除,指数相减.

例 1.18 计算:(a) 10×10^3, (b) $\frac{10^7}{10^3}$.

解 ☞ (a) $10\times10^3=10^{1+3}=10^4=10,000$

(b) $\frac{10^7}{10^3}=10^{7-3}=10^4=10,000$

1.10 对数运算

若 $n=c^b$,那么数 b 就称为以 c 为底的 n 的**对数**(其中 $c\neq1$ 且 $c>0$),记作 $b=\log_c n$.

例 1.19 求(a) 若 $4=2^b$,计算 $\log_2 4$, (b) 若 $\log_{10} n=2$,计算 n.

解 ☞ (a) 若 $4=2^b$,那么 $b=2$,于是 $\log_2 4=2$

(b) 若 $\log_{10} n=2$,那么 $n=10^2=100$.

通常,若 $\log_c n=b$,则 b 的**反对数**是 n.

例 1.20 根据下列式子写出反对数:(a) $\log_2 16=4$, (b) $\log_{10}10=1$.

解 ☞ (a) 4 的反对数是 16

(b) 1 的反对数是 10

两个正数的积的对数,等于这两个数的对数的和. 两个正数的商的对数,等于被除数的对数减去除数的对数的差. 一个正数的幂的对数,等于幂的底数的对数乘以幂指数.

例 1.21 计算下列各数的对数:(a) bc, (b) b/c, (c) a^b.

解 ☞ (a) $\log(bc)=\log b+\log c$

(b) $\log(b/c)=\log b-\log c$

(c) $\log(a^b)=b(\log a)$

1.11 代数表达式

代数表达式是指用四种基本运算符号[(+),(-),(×),(÷)]把**算术数**(有指定的数值)和**一般数**(表示数值的字母)连接而成的式子. **代数表达式的项**是指用+或-号分开的数、字母,或者用+或-号分开的它们的乘积或商. 只有一项的代数表达式称为**单项式**,有两项的代数表达式称为**二项式**,通常有两项或两项以上的代数表达式称为**多项式**.

例 1.22 指出下列代数表达式的项：$14a^2+10b-3c^4$.

解 ☞ 该代数表达式的项为：$(14a^2)$，$(10b)$，$(3c^4)$.

1.12 方程和公式

方程是表示两个代数式相等关系的式子. 例如：$a-b=c$；$\frac{15}{5}=3$；$y+3=4$；$\log x+y=2$. 每一个方程都由**等号**(=)和等号两边的表达式组成. **公式**是用代数符号表示定理或规则的方程. 例如：$c=\pi d$[圆的周长 c=它的直径 d×常数 π(π=3.14159…)]. 方程两边若进行相同运算(被 0 除除外)则得到等价方程.

例 1.23 写出下列方程进行指定运算的等价方程：(a) $a+b=c$，方程两边同时加 b，(b) $a+b=c$，方程两边同时平方.

解 ☞ (a) $a+2b=c+b$

(b) $(a+b)^2=c^2$，或 $a^2+2ab+b^2=c^2$

恒等方程是指等号两边的值恒等的方程. 例如$(y-3)(y-1)=y^2-4y+3$ 和 $7+2=9$，都是恒等方程. 这类方程也称为**恒等式**，通常用符号≡代替=表示. 包含字母的恒等式，对字母取任何值均成立.

例 1.24 通过取 $y=5$，说明$(y-3)(y-1)=y^2-4y-3$ 是恒等式.

解 ☞
$$(5-3)(5-1)=(5)^2-4(5)+3$$
$$8=8$$

恒等式对于字母取任意值均成立，而**条件方程**只对某些数值是成立的.

例 1.25 找出使条件方程 $2y+4=10$ 成立的 y 值.

解 ☞ 只有 $y=3$ 时，$2y+4=10$ 成立.

解方程是指找出使方程两边的值相等的一般数(字母)的值，这样的值称为**方程的解**. 对于恒等式$(y-3)(y-1)=y^2-4y+3$，y 的任意取值都是方程的解. 对于条件方程 $2y+4=10$，它只有一个解：$y=3$.

例 1.26 求解方程：$\frac{x}{10}+18=x$.

解 ☞
$$\frac{x}{10}+18=x$$
$$\frac{x}{10}-x=-18$$
$$x\left(\frac{1}{10}-1\right)=-18$$
$$x(-0.9)=-18$$
$$x=20$$

1.13 变量

在某一过程中可以取不同数值的一般数(见 1.11 节)称为**变量**. 因此，对于公式 $c=\pi d$(见 1.12 节)，如果问题要求计算直径分别为 1 in，2 in 和 3 in(1in(英寸)=2.540cm)的三个圆的周长，那么在这个问题中 c 和 d 都是变化的，因此它们都是变量.

在过程中保持同一数值的量称为**常量**. 常量分为**绝对常量**和**任意常量**两种. 绝对常量是指任何情况下都取相同值的量，例如算术数(例如 5，$\frac{1}{2}$，100)或取固定值的一般数[例如 e(自然对数的底，见习题 1.23)，π(见 1.12 节)]. 任意常量是指在某个过程中保持不变的一般数，但过程不同其取值可能会改变.

本书到这为止都是采用小写字母来表示方程或公式中的一般数. 如果没有特别指出，对于

表示变量的一般数,将用字母表中后几个字母(Z,Y,X 等等)的大写表示,而对于表示常量的一般数,用字母表中前几个字母(a,b,c 等等)的小写来表示.

1.14 单变量方程和二次公式

单变量线性方程,也称为**一元一次方程**,只含有常量和一个变量,并且变量的次数为 1. 例如 $2X-3=2$ 和 $aY-b=cY$. **单变量二次方程**也称为**一元二次方程**,含有常量和一个变量,并且变量的最高次数为 2. 例如 $aX^2+bX=c$,$aX^2=c$ 和 $3X^2-5X=0$. 如果方程中既有一次项又有二次项(例如 $aX^2+bX=c$),则称方程为**完全二次方程**;如果方程中只含有二次项(例如 $aX^2=c$),则称方程为**不完全二次方程**. 对于完全二次方程 $aX^2+bX+c=0$,它的解是

$$X=\frac{-b\pm\sqrt{b^2-4ac}}{2a} \tag{1.1}$$

上式称为**二次求根公式**. 对于任一二次项系数不为零的一元二次方程都可以用该公式求解(该公式的推导见习题 1.31).

例 1.27 用二次求根公式求解方程:$2X^2=-3X+9$.

解 先将方程化为一般的二次形式:$2X^2+3X-9=0$,于是 $c=-9$,$b=3$,$a=2$

$$X=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

$$=\frac{-3\pm\sqrt{3^2-[4\times(-9)\times 2]}}{2\times 2}$$

$$=\frac{-3\pm\sqrt{9-(-72)}}{4}=\frac{-3\pm\sqrt{81}}{4}=\frac{-3\pm 9}{4}$$

因此,

$$X=\frac{6}{4}=\frac{3}{2},\ X=\frac{-12}{4}=-3$$

1.15 统计中的变量

在统计中,变量是指所研究对象(人,物体,地点等)的特征,这些特征是可以测量的,并且对于一组对象中的不同对象可以取不同的值. 例如,如果我们研究一群儿童,那么由于儿童的体重是可以测量的,并且对于不同的儿童它可以取不同的值,所以儿童的体重是一个变量. 或者,如果我们研究一组马铃薯苗,那么我们可以考虑测量如下变量:马铃薯苗的高度、宽度、叶子的数量和马铃薯的个数. 变量的测量问题将在第二章中讨论.

变量的每一个测量值称为**观测值**. 在这本书中,用大写字母表示变量,而用同一字母的小写表示它的观测值(例如 X 表示变量,x 表示它的观测值).

观测值用数值表示的变量称为**定量变量**. 定量变量的观测值可以根据特征的数量(更重,更高,更富有等等)进行排序. 例如儿童的体重是一个定量变量,一个儿童体重为 45.2 磅(1 磅 =0.4536 千克),另一个更重的儿童体重为 50.9 磅. 类似的,马铃薯的个数也是一个定量变量.

如果变量的观测值只是一些不同的类别,不能按照其量值进行排序,则称该变量为**定性变量**. 例如树的种类就是一个定性变量,其观测值有:松树、枫树、白杨树以及山胡桃等.

1.16 可观测变量,假设变量和测量变量

统计中的变量可以分为**直接测量变量**和**间接测量变量**. 为理解这样的分类,考虑下面的例子.

遗传学家为了研究奥林匹克男子长跑冠军与短跑冠军之间的遗传差别,可以测量如下变量:身高、腿长、小腿和大腿的周长等.

在这个例子中,这些解剖学变量是直接测量变量,而遗传学变量(特定基因或基因组之间

的差别)是间接测量变量.直接测量变量又称为**可观测变量**,间接测量变量又称为**假设变量**(或称为**介入性变量**).

测量变量是可观测变量,是被测量事物的直接可测量特性,可以用具体测量尺度上的值来表示.测量变量可以是定性变量也可以是定量变量.以英尺测量的高度、以克测量的重量、一棵植物上叶子的数量以及花的种类等都是测量变量.

1.17　函数和关系

在讨论函数和关系之前,我们先定义两个概念:**集合**与**实数系**.

集合是指一些事物(物体、符号、数等)的全体,集合中的每一项称为集合的**元素**(或**成员**).

实数系是指由**有理数**(所有整数的比$\frac{a}{b}$,其中$b\neq 0$)和**无理数**(不能写成两个整数的比的数,如$\sqrt{2}$,π)构成的集合.

考虑两个变量X,Y,如果对于X在某一范围内的每一个确定的值,Y都有唯一确定的值y与它对应,那么就称Y是X的**函数**.函数通常有**定义域**(即X的取值范围)、**值域**(即Y的取值范围)和联系x值与y值的**对应规则**.

例 1.28　指出函数$Y=X^2$的定义域、值域和对应规则.

解 ☞　对于函数$Y=X^2$,X可以取任一实数,因此函数的定义域是实数系.对每一个特定的x值,对应的y值是零或正数,因此函数的值域是非负实数集合.函数的对应规则是$Y=X^2$.图1-1通过一些具体的值举例说明了这些概念.

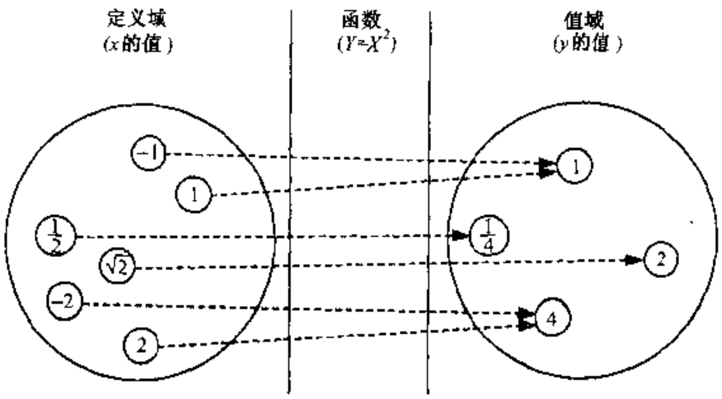


图 1-1

对于X在取值范围内的一个确定值,函数有唯一确定的对应值Y.因此,可以说Y值"依赖"于X值,所以我们称X为**自变量**,Y为**因变量**.

关系与函数的区别在于对应规则的不同.函数的对应规则是一对一,即一个X值对应一个且只能对应一个Y值;而关系的对应规则是一对多,即一个X值可以对应多个Y值.所以也称函数为**单值函数**,而称关系为**多值函数**.

例 1.29　指出关系$Y=X\pm 3$的定义域、值域和对应规则.

解 ☞　对于关系$Y=X\pm 3$,定义域和值域均为实数系,其对应规则表明每一个x都有两个y值相对应.图1-2举例说明了这一关系.

1.18　函数记号

函数可以写成变量间的关系式,如$Y=X^2$,或者写成以具体的变量值之间的联系.因此,$Y=X^2$可以用函数记号表示如下

$$y=f(x)=x^2$$

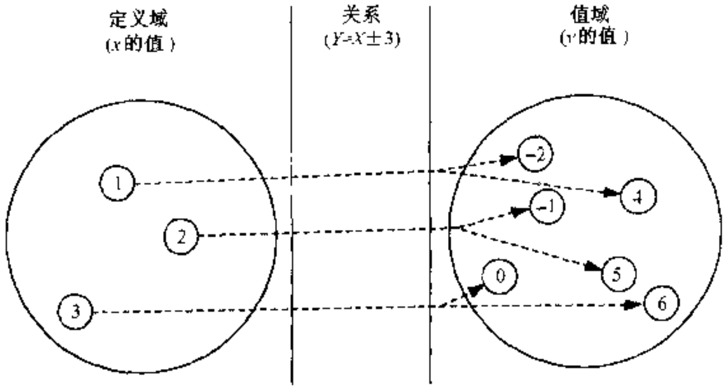


图 1-2

其中 $y=f(x)$ 读作 y 等于 x 的函数，而不能读成 y 等于 f 乘以 x. $f(x)$ 是一个一般的函数记号，我们也可以用 $F(z)$、$g(x)$、$h(y)$ 等其他记号来表示函数.

例 1.30 对函数 $y=f(x)=-3+2x+x^2$，计算(a) $f(0)$，(b) $f(1)$.

解 ☞ (a) $f(0)=-3+2(0)+0^2=-3-0+0=-3$

(b) $f(1)=-3+2(1)+1^2=-3+2+1=0$

1.19 统计中的函数

所有研究的根本目标都是探讨**因果关系**，找出**导致**事件发生的原因. 例如，植物学家想研究影响植物生长(果)的土壤的特性(因)，经济学家想确定影响汽车销售(果)的广告因素(因).

为了研究因果关系，研究人员使用统计技术确定自变量和因变量之间的函数关系. 在研究问题中，自变量是与"果"相联系的一个测量变量，因变量则是与"因"相联系的一个测量变量. 因变量的值在某种程度上是依赖于自变量的值. 因此，在上面的植物例子中，植物学家想说明植物高度(因变量)是土壤中氮含量(自变量)的函数；而在汽车销售例子中，经济学家想调查一个汽车公司过去每十年中销售的汽车数量(因变量)是否是该公司过去每十年中广告费用总数(自变量)的函数.

例 1.31 指出下列试验中的自变量和因变量

为了确定水的温度对鲑鱼生长的影响作用，我们饲养两组鲑鱼(每组 10 条)，一组水温控制在 20℃，另一组水温则控制在 24℃，其他条件都相同. 在孵化后 200 天，称一下两组中每一条鲑鱼的重量.

解 ☞ 在试验中，通过自变量的变化来研究其对因变量的影响. 这个试验中是通过水温的变化来研究其对鲑鱼的体重变化的影响，所以水温是自变量而鲑鱼的体重是因变量.

统计中的函数只需要定义域、值域和两个变量间的对应关系，而并不需要写成等式表达式. 例如，一个函数的定义域是一组男生的名字，值域是他们头发的颜色，对应规则为对每一个名字有且仅有唯一的头发颜色相对应.

1.20 实数轴和直角笛卡儿坐标系

实数轴(也称为**数轴**或**实轴**)，是用来表示实数系中所有实数的一条直线. 实数系(见 1.17 节)中的每一个数都可用实数轴上的一个点来表示.

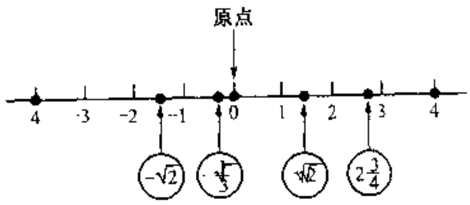


图 1-3

例 1.32 在实数轴上画出下列各数：-4，$-\sqrt{2}$，$-\frac{1}{3}$，0，$\sqrt{2}$，$2\frac{3}{4}$，4.

解 ☞　如图 1-3 所示.

直角笛卡儿坐标系(也称为**直角坐标系**)是由平面内有公共**原点**而且互相垂直的两条数轴构成的,如图 1-4 所示.横轴,也称为 **X 轴**,取向右方向为正方向.纵轴,也称为 **Y 轴**,取向上方向为正方向.X 轴和 Y 轴把平面(**XY 平面**)分成四个部分:**第一象限**、**第二象限**、**第三象限**和**第四象限**.在平面内建立了直角坐标系后,对于平面内的任意一点,都有一对有序实数(x,y)和它对应.第一个数,称为 **x 坐标或横坐标**,是点到 Y 轴的水平距离.第二个数,称为 **y 坐标或纵坐标**,是点到 X 轴的垂直距离.

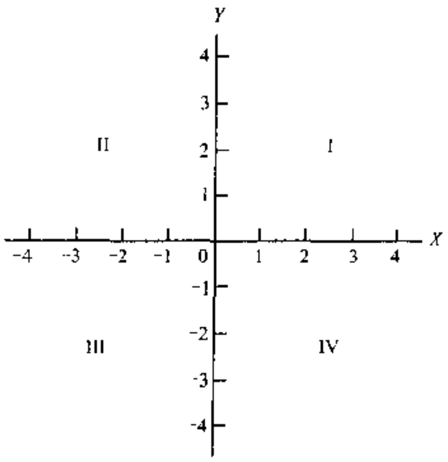


图 1-4

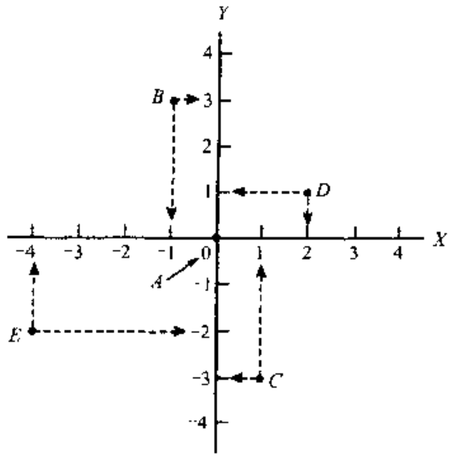


图 1-5

例 1.33　在直角坐标系中画出下列各点:$A(0,0)$;$B(-1,3)$;$C(1,-3)$;$D(2,1)$;$E(-4,-2)$.

解 ☞　这些点在直角坐标系中的位置如图 1-5 所示,其中虚线表示点到 X 轴和 Y 轴的距离.

1.21　函数的图象

函数的**图象**是函数中变量间关系的图形表示.把自变量 x 的值和函数 y 的对应值分别作为点的横坐标和纵坐标,可以在直角坐标系内绘出函数的图象.函数可以像方程(见 1.14 节)那样分类,因此,函数 $y=f(x)=c+bx$ 是一个线性或一次函数,它的图象是一条直线.

例 1.34　在直角坐标系内画出函数 $y=f(x)=4+2x$ 的图象.

解 ☞　该函数的图象如图 1-6 所示.

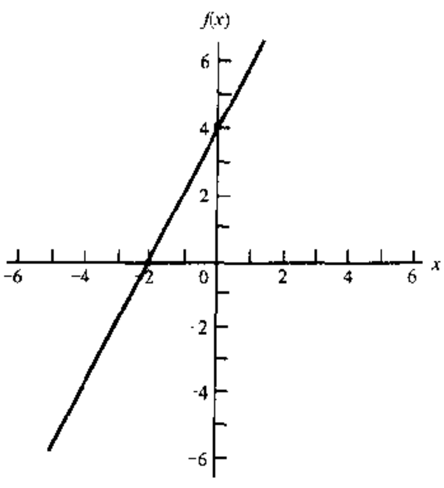


图 1-6

我们把形如 $y=f(x)=c+bx-ax^2$(其中 a,b,c 是常数,且 $a\neq0$)的函数叫做**二次函数**.一个二次函数的图象是一条**抛物线**,这条抛物线关于一条与 Y 轴平行的直线对称,这条直线称为**对称轴**.抛物线如果开口向下,则有**最大值**;如果开口向上,则有**最小值**.二次函数的图象的构造方法见习题 1.40.

1.22 序列、级数和求和符号

定义域为相继正整数的函数称为**序列**.如果定义域包括所有正整数,此序列称为**无限序列**.例如:$f(i)=i+1,i=1,2,3,\cdots,\infty$.这个序列的第一个数是 $f(1)=1+1=2$;第二个数是 $f(2)=2+1=3$;第三个数是 $f(3)=3+1=4$;等等.这个序列为:$2,3,4,\cdots,\infty$.序列中的每一个数称为序列的项.如果定义域只包括部分正整数,此序列称为**有限序列**.例如:$f(i)=x_i,i=1,2,3$.其中 x_i 中的 i 称为**下标**.这个序列中只有三项:x_1,x_2,x_3.

例 1.35 写出序列的各项:$f(i)=i^2-3,i=2,3,4$.

解 这个序列有三项:$f(2)=2^2-3=1$;$f(3)=3^2-3=6$;$f(4)=4^2-3=13$.

级数是序列的各项之和.对于无限序列 $f(i)=i+1,i=1,2,3,\cdots,\infty$,其级数是

$$2+3+4+\cdots+\infty$$

对于有限序列 $f(i)=x_i,i=1,2,3$,其级数是

$$x_1+x_2+x_3$$

符号 $\sum_{i=1}^{n}x_i$ 称为**求和符号**,它是级数 $x_1+x_2+x_3+\cdots+x_n$ 的记号.符号 $\sum$ 是希腊字母 σ 的大写形式,表示对它的右边序列求和.$\sum$ 下面的字母称作**求和下标**,或**求和变量**,它的数值表示序列定义域的下限.$\sum$ 上面的数值表示序列定义域的上限.对于 $\sum_{i=1}^{n}$,其上下限表示序列从 $i=1$ 加到 $i=n$(从 x_1 到 x_n).一般,求和下标用小写字母 i,j,k 表示.

例 1.36 求和:$\sum_{i=1}^{4}i^2$.

解 $\sum_{i=1}^{4}i^2=(1)^2+(2)^2+(3)^2+(4)^2=1+4+9+16=30$

如果 x_i 代表变量 X 的测量值,那么在统计学中 n 个测量值的全集是从 x_1 加到 x_n,这一级数记为 $\sum_{i=1}^{n}x_i$.通常我们省去上下限,直接记为 $\sum x_i$ 或 $\sum x$.

例 1.37 五个三年级的男孩的身高形成如下序列(1ft=0.3048 米):$x_1=2.1$ ft,$x_2=2.0$ ft,$x_3=1.9$ ft,$x_4=2.0$ ft,$x_5=1.8$ ft.求和:$\sum x_i$.

解

$$\begin{aligned}\sum x_i=\sum_{i=1}^{n}x_i=\sum_{i=1}^{5}x_i&=x_1+x_2+x_3+x_4+x_5\\&=2.1\text{ ft}+2.0\text{ ft}+1.9\text{ ft}+2.0\text{ft}+1.8\text{ft}\\&=9.8\text{ ft}\end{aligned}$$

1.23 不等式

符号($<,>,\leq,\geq$)称为**不等号**.用这些符号表示两个代数表达式**不相等关系**(小于,大于,小于等于,大于等于)的式子叫做**不等式**.不等号的箭头总是指向较小的表达式.

例 1.38 解释下列数学表达式:(a) $3<4$,(b) $5>2$,(c) $b>a$,(d) $b\geq a$,(e) $b\leq a$.

解 (a) $3<4$ 是指 3 小于 4

(b) $5>2$ 是指 5 大于 2

(c) $b>a$ 是指 b 大于 a

(d) $b\geq a$ 是指 b 大于或等于 a

(e) $b \leq a$ 是指 b 小于或等于 a

例 1.39　用不等号连接下列各组数：(a) 1,2，(b) −1，−2，(c) −1,2，(d) 1，−2.

解 ☞　(a) $1<2$，(b) $-1>-2$，(c) $-1<2$，(d) $1>-2$.

不等式的两边都加上或减去同一个数，不等号的方向不变. 不等式的两边都乘以或除以同一个正数，不等号的方向不变. 不等式的两边都乘以或除以同一个负数，不等号的方向改变.

例 1.40　对于不等式 $8>6$，(a) 两边都加上 5；(b) 两边都乘以 3；(c) 两边都乘以 −3.

解 ☞　(a) $8+5>6+5$，或 $13>11$

(b) $8\times3>6\times3$，或 $24>18$

(c) $8\times(-3)<6\times(-3)$，或 $-24<-18$

对于单变量不等式，如果用一个数代替变量，不等式仍然成立，则称这个数为该**不等式的解**. 不等式的所有解构成不等式的**解集**.

例 1.41　解不等式：$X+7>-3$.

解 ☞　不等式两边都减去 7，得

$$X>-10$$

所以这个不等式的解集为大于 −10 的所有实数.

习题解答

分数运算

1.1　令 a，b 表示两个数，则下列哪些分数与 $\frac{a}{2b}$ 等价的：$\frac{a^2}{2ab}$，$\frac{2a}{2b}$，$\frac{1}{(2b)/(a)}$，$\frac{a-1}{2b-1}$？

解 ☞

$$\frac{a}{2b}=\frac{(a)(a)}{(a)(2b)}=\frac{a^2}{2ab}=\frac{(a)/(a)}{(2b)/(a)}=\frac{1}{(2b)/(a)}$$

而

$$\frac{a}{2b}\neq\frac{2a}{2b}$$

$$\frac{a}{2b}\neq\frac{a-1}{2b-1}$$

1.2　计算下列各式：(a) $\frac{g}{h}-\frac{a}{b}$，(b) $\frac{1}{2}\div1\frac{1}{2}$.

解 ☞　(a) $\frac{g}{h}-\frac{a}{b}=\frac{b\times g}{b\times h}-\frac{a\times h}{b\times h}=\frac{bg-ah}{bh}$

(b) $\frac{1}{2}\div1\frac{1}{2}=\frac{1}{2}\times\frac{2}{3}=\frac{2}{6}=\frac{1}{3}$

带符号数的运算

1.3　对下列符号数进行加减运算：(a) (−12)+(−3)+(−5)+(−6)，(b) 14−(−15)+(+2)−(+3).

解 ☞　(a) $(-12)+(-3)+(-5)+(-6)=-26$

(b) $14-(-5)+(+2)-(+3)=21-3=18$

1.4　对下列符号数进行乘法和除法运算：(a) (0.4)(−0.002)，(b) (−0.29)·(−0.36)，(c) (−0.009)÷(−0.03)，(d) (4.2)÷(−1.2).

解 ☞　(a) $(0.4)(-0.002)=-0.0008$

(b) $(-0.29)\cdot(-0.36)=0.1044$

(c) $(-0.009)\div(-0.03)=0.300$

(d) $(4.2)\div(-1.2)=-3.5$

1.5　确定运算顺序并计算：(a) 2÷10+9，(b) (0.004/0.002)+ (0.9/0.003).

解☞ (a) $2\div10+9=0.2+9=9.2$

(b) $(0.004/0.002)+(0.9/0.003)=(2)+(300)=302$

1.6 完成下列与零有关的运算:(a) $4+0+3$,(b) $4\times0\times3$,(c) $12/0$.

解☞ (a) $4+0+3=7$

(b) $4\times0\times3=0$

(c) 数被零除无意义.

舍入运算

1.7 对下列数进行取整运算:(a) 13.499990,(b) 13.50000.

解☞ (a) 13.499990 舍入到 13

(b) 13.50000 舍入到 14

1.8 舍入运算:(a) 40.195 舍入到二位小数,(b) 0.020936 舍入到三位小数.

解☞ (a) 40.195 舍入到 40.20

(b) 0.020936 舍入到 0.021

绝对值

1.9 证明 $|c|+|d|\neq|c+d|$

解☞ 令 $c=5,d=-4$,则

$$|5|+|-4|\neq|5+(-4)|$$
$$9\neq1$$

阶乘运算

1.10 计算下列阶乘:(a) $0!$,(b) $1!$.

解☞ (a) 由定义知 $0!=1$

(b) $1!=1$

1.11 计算:(a) $\dfrac{6!}{(4-2)!}$,(b) $\dfrac{3!}{(3!)(2!)}$.

解☞ (a) $\dfrac{6!}{(4-2)!}=\dfrac{6\times5\times4\times3\times2\times1}{2!}=\dfrac{6\times5\times4\times3}{1}=360$

(b) $\dfrac{3!}{(3!)(2!)}=\dfrac{1}{2!}=\dfrac{1}{2}$

开方和根

1.12 计算:$\sqrt[3]{125}$

解☞ $\sqrt[3]{125}=125$ 的立方根$=+5$

1.13 求解 $\sqrt[3]{-8}$ 的主 n 次方根:

解☞ $\sqrt[3]{-8}=-2$

平方根运算

1.14 完成下列指定运算:(a) $\sqrt{28}+\sqrt{63}$,(b) $5\sqrt{(2-1/25)}$.

解☞ (a) $\sqrt{28}+\sqrt{63}=(\sqrt{4}\sqrt{7})+(\sqrt{9}\sqrt{7})=2\sqrt{7}+3\sqrt{7}=5\sqrt{7}$

(b) $5\sqrt{(2-1/25)}=\sqrt{25}\sqrt{(2-1/25)}=\sqrt{25(2-1/25)}=\sqrt{50-1}=\sqrt{49}=7$

幂运算

1.15 将下列各式表示成根式:(a) $4^{1/2}$,(b) $8^{1/3}$.

解☞ (a) $4^{1/2}=\sqrt{4}=2$

(b) $8^{1/3}=\sqrt[3]{8}=2$

1.16 将下列各式表示成分数:(a) 7^{-2}, (b) $4^{-1/2}$.

解☞ (a) $7^{-2}=\frac{1}{7^2}=\frac{1}{49}$

(b) $4^{-1/2}=\frac{1}{\sqrt{4}}=\frac{1}{2}$

1.17 写出下列各式的值:$10^6, 10^4, 10^2, 10^0, 10^{-2}, 10^{-4}, 10^{-6}$.

解☞ $10^6=1,000,000$; $10^4=10,000$; $10^2=100$; $10^0=1$; $10^{-2}=0.01$; $10^{-4}=0.0001$; $10^{-6}=0.000001$

1.18 计算:(a) $10^3\times10^{-4}$, (b) $\frac{7^8}{7^{-3}}$.

解☞ (a) $10^3\times10^{-4}=10^{3-4}=10^{-1}=0.1$

(b) $\frac{7^8}{7^{-3}}=7^{8-(-3)}=7^{11}$

1.19 计算:(a) $(10^2)^2$, (b) $(10^{-2})^{-2}$, (c) $(ab)^n$.

解☞ (a) $(10^2)^2=10^{2\times2}=10^4=10,000$

(b) $(10^{-2})^{-2}=10^{(-2)(-2)}=10^4=10,000$

(c) $(ab)^n=(a^n)(b^n)$

1.20 将下列各式进行指数运算与开方运算之间的转换:(a) $10^{4/5}$, (b) $\sqrt[3]{5^6}$.

解☞ (a) $10^{4/5}=\sqrt[5]{10^4}$

(b) $\sqrt[3]{5^6}=5^{6/3}=5^2=25$

对数运算

1.21 若 $\log_c 1,000=3$,则 c 等于多少?

解☞ 如果 $\log_c 1,000=3$,那么 $c^3=1,000$,所以 $c=\sqrt[3]{1,000}=10$.

1.22 根据下列各式写出反对数:(a) $\log_a n=d$, (b) $\log_5 25=2$,

解☞ (a) d 的反对数等于 n

(b) 2 的反对数等于 25

1.23 用电子计算器计算:(a) 100 的常用对数,(b) 100 的自然对数

解☞ (a) 100 的常用对数等于 2

(b) 100 的自然对数等于 4.60517(精确到小数点后第 5 位)

1.24 用电子计算器求解:若 1.69897 是数 a 的常用对数,则它的反对数是多少?

解☞ 1.69897 的反对数$=10^{1.69897}=49.9999995$,或 50

1.25 用电子计算器求解:若 3.91202 是数 b 的常用对数,则它的反对数是多少?

解☞ 3.91202 的反对数$=e^{3.91202}=49.9998497$,或 50

1.26 计算下列各式的常用对数:(a) $4bc$,(b) $7/5$,(c) $1\frac{5}{27}$.

解☞ (a) $\log(4bc)=\log 4+\log b+\log c=0.60206+\log b+\log c$

(b) $\log(7/5)=\log 7-\log 5=0.84510-0.69897=0.14613$

(c) $\log(1\frac{5}{27})=\log(32/27)=\log 32-\log 27=1.50515-1.43136=0.07379$

1.27 计算下式的常用对数:$\sqrt[4]{\frac{(49)(27)}{(3.1)^3}}$

解☞ $\log\left(\sqrt[4]{\frac{(49)(27)}{(3.1)^3}}\right)=\log\left[\left(\frac{(49)(27)}{(3.1)^3}\right)^{\frac{1}{4}}\right]=\frac{1}{4}[(\log49+\log27)-3(\log3.1)]$

$$=\frac{1}{4}[(1.69020+1.43136)-3(0.49136)]$$

$$=\frac{1}{4}(3.12156-1.47408)$$

$$=0.41187$$

代数表达式

1.28 指出代数表达式的各项:$\frac{2\sigma}{a}-\mu$

解☞ 该代数表达式的各项为:$\left(\frac{2\sigma}{a}\right)$,$(\mu)$.

方程和公式

1.29 根据指定运算写出等价方程:(a) $a+b=c$,方程两边都减去 b;(b) $ab=c$,方程两边都除以 a;(c) $\frac{b}{a}=c$,方程两边都乘以 a.

解☞ (a) $a=c-b$

(b) $b=\frac{c}{a}$

(c) $b=ca$

1.30 求解方程:(a) $x+5=10$,(b) $(x-2)^2=4$.

解☞ (a) $x+5=10$

$$x=5$$

(b) $(x-2)^2=4$

$$(x-2)=2$$

$$x=4$$

(编者注:$(x-2)=\pm2$,$x_1=0$,$x_2=4$.原文只有一个根是错的)

单变量方程和二次公式

1.31 通过求解完全二次方程 $aX^2+bX+c=0$ 推导二次求根公式.

解☞ 求解完全二次方程并推导二次求根公式的步骤如下:

(1)方程两边都乘以 $\frac{1}{a}$:

$$\frac{1}{a}(aX^2+bX+c)=\frac{1}{a}(0)$$

$$X^2+\frac{b}{a}X+\frac{c}{a}=0$$

(2)方程两边都加上 $\left[-\left(\frac{c}{a}\right)+\left(\frac{b}{2a}\right)^2\right]$:

$$X^2+\frac{b}{a}X+\frac{c}{a}-\frac{c}{a}+\left(\frac{b}{2a}\right)^2=-\frac{c}{a}+\left(\frac{b}{2a}\right)^2$$

$$X^2+\frac{b}{a}X+\left(\frac{b}{2a}\right)^2=-\frac{c}{a}+\frac{b^2}{4a^2}$$

(3)整理方程的左边:

$$\left(X+\frac{b}{2a}\right)^2=-\frac{c}{a}+\frac{b^2}{4a^2}$$

(4)整理方程的右边:

$$\left(X+\frac{b}{2a}\right)^2=-\frac{4a\times c}{4a\times a}+\frac{b^2}{4a^2}=-\frac{4ac}{4a^2}+\frac{b^2}{4a^2}=\frac{b^2-4ac}{4a^2}$$

(5)方程两边都开平方：

$$X+\frac{b}{2a}=\pm\sqrt{\frac{b^2-4ac}{4a^2}}=\pm\frac{\sqrt{b^2-4ac}}{2a}$$

(6)得到 X 的解：

$$X=-\frac{b}{2a}\pm\frac{\sqrt{b^2-4ac}}{2a}=\frac{-b}{2a}\pm\sqrt{\frac{b^2-4ac}{2a}}$$

1.32　根据二次求根公式[方程(1.1)]求解：$X^2=12X-36$.

解 ☞

$$X^2-12X+36=0$$

因此 $a=1, b=-12, c=36$

$$X=\frac{12\pm\sqrt{12^2-(4\times36\times1)}}{2\times1}=\frac{12\pm\sqrt{144-144}}{2}=\frac{12}{2}=6$$

统计中的变量

1.33　颜色是定性变量还是定量变量？

解 ☞　物体的颜色特征可以被描述成定性或定量变量.颜色的物理学基础是光的波长，它可以表示为一个数[例如，深红色大约是 $8,000\times10^{-10}$ 米(或 8,000 埃)].如果用波长来描述颜色，那么颜色就是一个定量变量.然而，如果颜色变量被分成无序的类别，那么它就是定性的变量.例如一组人可以通过头发的颜色分成黑色、红色、金色、棕色、灰色和白色.

函数和关系

1.34　写出函数的定义域和值域：$Y=\frac{1}{X}$.

解 ☞　对于 $Y=\frac{1}{X}$，定义域和值域均为不为零的实数.

1.35　给定函数 $y=f(x)=2^x$，求：(a) $f(0)$，(b) $f(3)$，(c) $f(6)$.

解 ☞　(a) $f(0)=2^0=1$

(b) $f(3)=2^3=8$

(c) $f(6)=2^6=64$

1.36　下表给出了一函数的自变量 x 和因变量 y 的三组值，求函数 $f(x)$.

x	2	3	4
y	2	4	6

解 ☞

$$y=f(x)=-2+2x$$

1.37　已知函数 $y=f(x)=-5+5x^3$，请完成表格.

x	1	2	3
y			

解 ☞

x	1	2	3
y	0	35	130

实数轴和直角笛卡儿坐标系

1.38　写出图 1-7 中点的坐标.

解☞ 为了得到直角坐标系中点的坐标，过该点分别作 X 轴和 Y 轴的垂直线，这些垂直线与坐标轴的交点就是该点的坐标. 通过该方法得到这些点的坐标为：$A(1,2\frac{1}{2})$；$B(1,-1)$；$C(-2,-3)$；$D(-2,2)$.

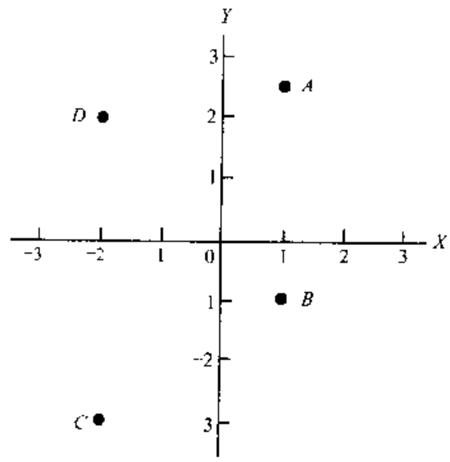


图 1-7

函数的图象

1.39 用函数 $y=f(x)=4+2x$ 的斜率和 Y 轴的截距在直角坐标系中作出它的图象.

解☞ 对于形式为 $y=f(x)=c-bx$ 的线性函数，其中 c 和 b 都是不为零的实数. c 是这一直线的 Y 轴上的截距，b 则是它的斜率. 函数 $y=f(x)=4+2x$ 的斜率为 2，它在 Y 轴上的截距为 4，也就是说它与 Y 轴相交于点(0,4)，由斜率和此交点可以得到另一个过直线的点为(1,6)，于是得到该函数的图象，如图 1-8 所示.

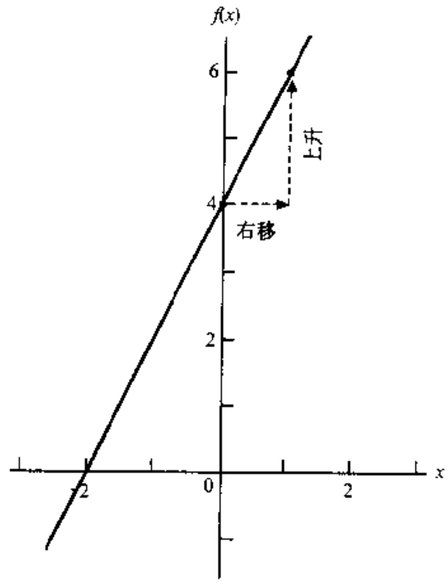


图 1-8

1.40 在直角坐标系中作出二次函数 $y=f(x)=-4+3x+x^2$ 的图象.

解☞ 为了得到这个二次函数与 X 轴的两个交点，我们令 $y=0$，得到：$0=-4+3x+x^2$. 由二次求根公式[方程(1.1)]得到该方程的解为：$x=1$，$x=-4$. 所以这个函数与 X 轴的两个交点为：(1,0)，(-4,0).

通过二次求根公式的如下形式

$$x=\frac{-b}{2a}\pm\frac{\sqrt{b^2-4ac}}{2a}$$

可以得到函数与对称轴交点的 X 轴坐标为$\frac{-b}{2a}$. 在这里，

$$\frac{-b}{2a}=\frac{-3}{2\times 1}=-1.5$$

通过令 $x=0$，得到：$y=-4+3(0)+(0)^2$. 所以该抛物线与 Y 轴交点的坐标为$(0,-4)$.

抛物线的开口向上还是向下取决于函数中 a 的符号：当 a 为正时，抛物线的开口向上；当 a 为负时，抛物线的开口向下. 在这个函数中，$a=1$，所以抛物线开口向上，并且有最小值. 这个最小值是该抛物线与其对称轴的唯一交点. 由前面已知，该交点的横坐标为-1.5，则其纵坐标为

$$y=-4+3(-1.5)+(-1.5)^2=-6.25$$

于是函数的最小值坐标为$(-1.5,-6.25)$.

根据函数与 X 轴和 Y 轴的交点、对称轴以及最小值，可以作出如图 1-9 所示的图象.

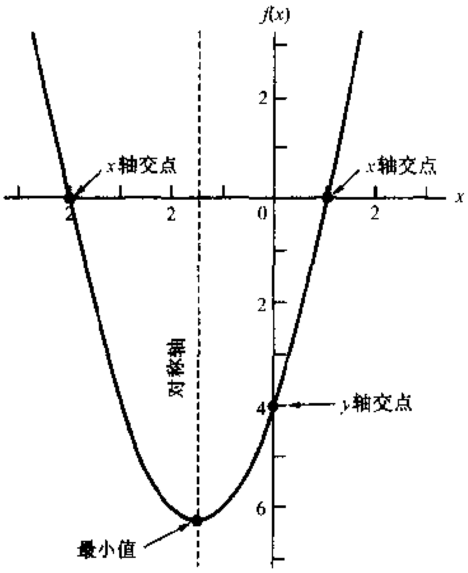


图 1-9

序列、级数和求和记号

1.41　求和：(a) $\sum_{i=3}^{6}i^2$，(b) $\sum_{i=4}^{7}(i-1)$.

解　(a) $\sum_{i=3}^{6}i^2=(3)^2+(4)^2+(5)^2+(6)^2=9+16+25+36=86$

(b) $\sum_{i=4}^{7}(i-1)=(4-1)+(5-1)+(6-1)+(7-1)=3+4+5+6=18$

1.42　求和：(a) $\sum_{i=1}^{5}6$，(b) $\sum_{i=2}^{5}6$.

解　(a) 通常，若 a 是常数，则 $\sum_{i=1}^{n}a=na$. 所以

$$\sum_{i=1}^{5}6=6+6+6+6+6=5\times 6=30$$

(b) 在这里，下限是 2 而不是 1，所以 $\sum_{i=2}^{n}a=(n-1)a$. 于是

$$\sum_{i=2}^{5}6=(5-1)\times 6=24$$

1.43　证明：$\sum_{i=1}^{3}3x_i=3\sum_{i=1}^{3}x_i$.

解 ☞ 通常,若 a 是常数,则 $\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i$,所以

$$\sum_{i=1}^{3} 3x_i = 3x_1 + 3x_2 + 3x_3 = 3(x_1 + x_2 + x_3) = 3\sum_{i=1}^{3} x_i$$

1.44 证明:

$$\sum_{i=1}^{3}\left(\frac{x_i}{2}\right) = \frac{\sum_{i=1}^{3} x_i}{2}$$

解 ☞ 通常,若 a 是常数,则 $\sum_{i=1}^{n}\left(\frac{x_i}{a}\right) = \frac{\sum_{i=1}^{n} x_i}{a}$,所以

$$\sum_{i=1}^{3}\left(\frac{x_i}{2}\right) = \frac{x_1}{2} + \frac{x_2}{2} + \frac{x_3}{2} = \frac{x_1 + x_2 + x_3}{2} = \frac{\sum_{i=1}^{3} x_i}{2}$$

1.45 证明:

$$\sum_{i=1}^{4}(4 + \pi + x_i) = 16 + 4\pi + \sum_{i=1}^{4} x_i$$

解 ☞ 通常,若 a,b 是常数,则 $\sum_{i=1}^{n}(a + b + x_i) = na + nb + \sum_{i=1}^{n} x_i$,所以

$$\sum_{i=1}^{4}(4 + \pi + x_i) = (4 \times 4) + (4 \times \pi) + \sum_{i=1}^{4} x_i = 16 + 4\pi + \sum_{i=1}^{4} x_i$$

1.46 证明:

$$\sum_{i=1}^{3}(3x_i + 2y_i) = 3\sum_{i=1}^{3} x_i + 2\sum_{i=1}^{3} y_i$$

解 ☞ 通常,若 a,b 是常数,则

$$\sum_{i=1}^{n}(ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i$$

所以

$$\sum_{i=1}^{3}(3x_i + 2y_i) = 3\sum_{i=1}^{3} x_i + 2\sum_{i=1}^{3} y_i$$

不等式

1.47 对于不等式 $6<7$,(a)两边都加上 8;(b)两边都减去 8.

解 ☞ (a) $6+8<7+8$,或 $14<15$

(b) $6-8<7-8$,或 $-2<-1$

1.48 对于不等式 $12>8$,(a)两边都乘以 2;(b)两边都除以 2.

解 ☞ (a) $12\times2>8\times2$,或 $24>16$

(b) $12/2>8/2$,或 $6>4$

1.49 对于不等式 $2>1$,(a)两边都乘以 -1;(b)两边都除以 -1.

解 ☞ (a) $2\times(-1)<1\times(-1)$,或 $-2<-1$

(b) $2/(-1)<1/(-1)$,或 $-2<-1$

1.50 解不等式:$-3X-2<2X+5$.

解 ☞ 求解步骤:

(1) 不等式两边都加上 $3X$,得

$$-2 < 5X + 5$$

(2) 不等式两边都减去 5,得

$$-7 < 5X$$

(3) 不等式两边都除以 5,得

$$\frac{-7}{5} < X$$

所以这个不等式的解为大于$\frac{-7}{5}$的所有实数.

1.51　解不等式：$\frac{X}{2}+1 \leq X-1$.

解　求解步骤：

(1) 不等式两边都加上1，得

$$\frac{X}{2}+2 \leq X$$

(2) 不等式两边都减去$\frac{X}{2}$，得

$$2 \leq X-\frac{X}{2}$$

(3) 化简后，不等式两边都乘以2，得

$$4 \leq X$$

所以这个不等式的解为大于等于4的所有实数.

1.52　解释下列单变量不等式：(a) $-9 \leq X < 2$，(b) $-2 < X \leq -1$.

解　(a) $-9 \leq X < 2$ 说明变量 X 只能取大于等于-9且小于2的值. 这表明不等式的解集包含-9到2之间的所有实数，并且包括-9但不包括2.

(b) $-2 < X \leq -1$ 说明变量 X 只能取大于-2且小于等于-1的值. 这表明不等式的解集包含-2到-1之间的所有实数，并且包括-1但不包括-2.

1.53　指出下列单变量不等式的错误：$-7 > X > 7$

解　大于7而小于-7的数是不存在的，所以这个不等式是无效的.

补充习题

分数运算

1.54　下列哪些分数与$\frac{4\div 3}{2\times 3}$等价：$\frac{2}{9}$，$\frac{4/2}{27\div 3}$，$\frac{(2)(2)}{36/2}$，$\frac{3\cdot 6}{6\frac{1}{2}}$.

答案　$\frac{4\div 3}{2\times 3}=\frac{2}{9}=\frac{4/2}{27\div 3}=\frac{(2)(2)}{36/2}$，

而$\frac{4\div 3}{2\times 3}\neq\frac{3\cdot 6}{6\frac{1}{2}}$.

1.55　计算下列各式：(a) $\frac{8\div 3}{(2)(3)}-\frac{7/2}{6\times 4}$，(b) $\frac{(14)(3)}{19/2}\div\frac{18\times 4}{16\div 3}$.

答案　(a) $\frac{43}{144}$，(b) $\frac{56}{171}$

带符号数的运算

1.56　对下列符号数进行加减运算：(a) $1.3+(-1.7)-(-2.3)-(+4.2)-(-3.1)$，(b) $(-0.93)+(-0.26)\ \ (-3.91)+(-2.1)$.

答案　(a) 0.8，(b) 0.62

1.57　对下列符号数进行乘法和除法运算：(a) $(1,800)\div(-0.2)$，(b) $(-3.63)(-0.0001)$，(c) $(-0.0004)/(-0.002)$.

答案　(a) $-9,000.0$，(b) 0.000363，(c) 0.2000

1.58　确定运算顺序并计算：$[(-3)\times(-4)+(2)]\div[(2)-(-0.5)]$.

答案　5.6

1.59　完成下列与零有关的运算：(a) $4\times 0+3$，(b) $0/12+3$.

答案 (a) 3, (b)3

舍入运算

1.60 对下列数进行取整运算:(a) 24.501、(b) 24.50.

答案 (a) 25、(b) 24

1.61 舍入运算:(a) 2.125 舍入到二位小数, (b) 4.93250 舍入到三位小数.

答案 (a) 2.12、(b) 4.932

绝对值

1.62 计算下列各数的绝对值:(a) $10/-2$、(b) 0, (c) $7-10$.

答案 (a) 5, (b) 0 , (c) 3

阶乘运算

1.63 计算:$\frac{6!}{5(3!)}$.

答案 24

开方和根

1.64 计算:$\sqrt[4]{1}$.

答案 有两个可能答案:+1 和 −1.按照惯例$\sqrt[4]{1}=+1$, $-\sqrt[4]{1}=-1$.

1.65 求解 $\sqrt[4]{625}$的主 n 次方根.

答案 按照惯例 $\sqrt[4]{625}$的主 4 次方根是 5.

平方根运算

1.66 完成下列指定运算:(a) $\sqrt{27}-\sqrt{12}$, (b) $\frac{3\sqrt{45}-\sqrt{20}}{\sqrt{125}/5}$.

答案 (a) $\sqrt{3}$, (b) 7

幂运算

1.67 将下列各式表示成分数:(a) 8^{-3}, (b) $8^{-1/3}$.

答案 (a) $\frac{1}{512}$、(b) $\frac{1}{2}$

1.68 用科学记数法表示下列各数:(a) 0.0237, (b) 0.11627, (c) 11,627.

答案 (a) 2.37×10^{-2}, (b) 1.1627×10^{-1}, (c) 1.1627×10^{4}

1.69 计算:(a) $(3\times7^{3})(2\times7^{-2})$, (b) $\frac{6\times10^{-3}}{3\times10^{-1}}$.

答案 (a) 42、(b) 0.02

1.70 计算:(a) $(3\times5)^{3}$、(b) $(3\times5)^{-2}$.

答案 (a) 3,375、(b) 0.00444

1.71 将下列各式进行指数运算与开方运算之间的转换:(a) $9^{4/3}$、(b) $\sqrt[5]{9^{2}}$.

答案 (a) $\sqrt[3]{9^{4}}$、(b) $9^{2/5}$

对数运算

1.72 若 $\log_c 0.001=-3$,则 c 等于多少?

答案 10

1.73 若 3.17609 是数 a 的常用对数,则它的反对数是多少?

答案 1,500

1.74 若 −1.89712 是数 b 的自然对数,则它的反对数是多少?

答案　0.15

1.75　计算下列各式的常用对数:(a) $\frac{5}{7}$, (b) 111^{12}.

答案　(a) -0.14613, (b) 24.54388

1.76　计算$\left(\frac{(15)^4\sqrt{22}}{(15)^{1/3}}\right)^{-1/3}$的常用对数.

答案　-1.26915

代数表达式

1.77　指出代数表达式 $14+2(a+b)$的各项.

答案　(14), (2a), (2b)

方程和公式

1.78　求解方程:$2y^2=8$.

答案　$y=2$ 或 $y=-2$

1.79　求解 c 的方程:(a) $a=b(1+c)$, (b) $c^2a-b=2b$.

答案　(a) $c=\frac{a}{b}-1$, (b) $c=\sqrt{\frac{3b}{a}}$

单变量方程和二次公式

1.80　根据二次求根公式[方程(1.1)]求解:$4X^2=1$.

答案　$X=\frac{0\pm\sqrt{16}}{8}=\pm\frac{4}{8}=\pm\frac{1}{2}$

函数和关系

1.81　写出函数 $Y=\sqrt{X-2}$ 的定义域和值域

答案　定义域是大于等于 2 的所有实数,值域是非负实数集.

1.82　给定函数 $y=f(x)=7$,求:(a) $f(0)$, (b) $f(3)$, (c) $f(6)$.

答案　无论 x 取什么值,均有 $y=f(x)=7$.

实数轴和直角笛卡儿坐标系

1.83　写出图 1-7 中点 A 和点 B 的横坐标和纵坐标.

答案　A 的横坐标为 1,纵坐标为 2.5;B 的横坐标为 1,纵坐标为 -1.

函数的图象

1.84　用函数 $y=f(x)=3-0.5x$ 的斜率和 Y 轴的截距在直角坐标系中作出它的图象.

答案　函数 $y=f(x)=3-0.5x$ 的斜率为 -0.5,它在 Y 轴上的截距为 3,也就是说它与 Y 轴相交于点 (0,3),该函数的图象如图 1-10 所示.

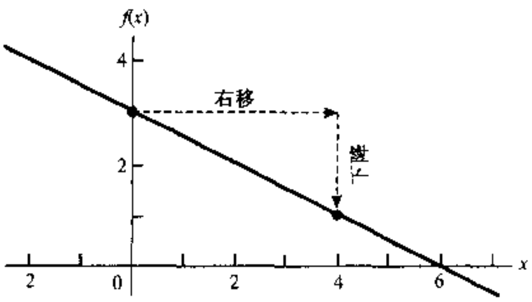


图 1-10

序列、级数和求和记号

1.85　求和:(a) $\sum_{i=3}^{6} i-1$,(b) $\sum_{i=1}^{5} x_i^2$.

答案　(a) 17,(b) $x_1^2+x_2^2+x_3^2+x_4^2+x_5^2$

1.86　对例 1.37 中的高度,计算:(a) $\sum x_i^2$,(b) $(\sum x_i)^2$.

答案　(a) 19.26 ft²,(b) 96.04 ft²

1.87　计算:$\sum_{i=2}^{4}(5+3i+i^2)$.

答案　71

1.88　计算:$\sum_{i=1}^{3}(\frac{i^2}{4}+3i^2)$.

答案　45.5

不等式

1.89　对于不等式 $b \geq a$,(a)两边都加上 5;(b)两边都减去 5.

答案　(a) $b+5 \geq a+5$,(b) $b-5 \geq a-5$

1.90　对于不等式 $3<4$,(a)两边都乘以 3;(b)两边都除以 2.

答案　(a) $9<12$,(b) $1\frac{1}{2}<2$

1.91　对于不等式 $6<7$,(a)两边都乘以 -2;(b)两边都除以 -7.

答案　(a) $-12>-14$,(b) $-\frac{6}{7}>-1$

第二章　统计资料的特征

2.1　测量尺度

测量尺度是指一种可应用于可观测变量从而产生测量变量(见1.16节)的工具.当一个以英尺为单位的刻度尺用于测量物体长度时,它就是一个测量尺度,它能够产生以英尺为单位的长度测量变量.运用这样一个测量尺度,每一个被测量物体就与尺度上的一个具体值相对应.

产生定性测量变量(见1.15节)的测量尺度是将可观测变量分到一系列唯一的、无序的类别的集合.例如,可观测变量是头发颜色,那么测量尺度应包括这些类别:黑色、红色、金色、棕色、灰色和白色.这样的尺度应含有足够的类别使得要测量的每一事物都能分到一个并且是唯一的类别中去.

产生定量测量变量(见1.15节)的测量尺度也是由一系列唯一的类别组成,但是这些类别可以由大到小进行排序.通常,这些类别是连续递增的数值.例如,要测量一群人的高度,我们可以使用以0.1厘米递增的尺度,那么每一个人的身高就与该尺度上的一个并且是唯一的类别相对应.

2.2　测量的操作定义

测量是将一个测量尺度用于可观测变量从而产生测量变量的过程.它包括将每一个被测量事物与测量尺度中的一个类别相对应的规则.**测量的操作定义**给出了进行测量的具体步骤或**操作**,即将测量尺度运用到可观测变量上去的步骤.这一定义应该是充分精确并且详细的,使得每一个执行测量过程的人都能得到基本上相同的测量结果.

例2.1　用两个米尺、胶带和一个梯凳测量某班级所有同学的身高,精确到毫米.请写出测量的操作定义.

解 ☞　测量步骤如下:

(1)选择一个门框,用胶带将两个米尺固定在门框上(一个米尺在另一米尺的上方,上方米尺的底端刻度线与下方米尺的顶端刻度线平齐),以产生一个垂直的两米尺度.

(2)班上每一个同学依次脱鞋,背靠门框站直.

(3)测量者站在梯凳上,眼睛平视每个同学的头顶,从两米尺度上读出他们的身高,精确到毫米.

2.3　测量水平和测量单位

测量尺度共有四个水平:**名义尺度**(见2.4节)、**次序尺度**(见2.5节)、**区间尺度**(见2.6节)和**比例尺度**(见2.7节).名义尺度得到定性测量变量,次序尺度、区间尺度和比例尺度得到定量测量变量.

除了次序测量尺度,所有用来产生定量测量变量的尺度都具有统一的、标准的**测量单位**.这些测量单位不仅与观测变量的类型(例如长度、质量、时间、温度)一致,而且给出了测量尺度的距离,以作为测量值之间进行比较的标准参照系.本书中的两个基本单位系统是**英制单位系统**(例如英尺、英磅、秒)和**米制单位系统**(或**国际单位系统**,例如米、克、秒).

2.4　名义水平测量

名义水平测量是最基本的度量水平,它将事物分到唯一的类中.这些类必须是**互斥**(没有事物被分到一个以上的类中)而且是**完备**的(每一个事物至少被分到一个类中),使被测量的事物能分到且只能分到一个类别中.在数学上,这一性质可以用等于和不等于符号($=,\neq$)表示.

名义尺度划分的类不能排序(例如,从小到大),数值只是用作标记这些类.例如,汽车牌照就是一种名义尺度.名义尺度最少包含两个类别(如硬币正面朝上或反面朝上),最多可以根据需要确定.名义尺度的其他一些例子如:鱼的种类(例如,鲨鱼、比目鱼,鲑鱼)、生病或不生病等.

2.5 次序水平测量

次序水平测量高于名义水平测量.它不仅保留了名义尺度的性质(事物能分到且只能分到一个类中,即=,≠).并且这些类可以根据被测量的特征的等级进行排序.每一类都可以与其相邻类进行比较,来判断是大于(>)还是小于(<).例如,把物体的大小分成大、中、小三类,分别用1、2、3表示;把电影质量由坏到好分成5类,分别用数字1到5表示;把儿童在游戏中的犯规分成十个数字表示的等级.

尽管次序尺度得到的是定量测量变量,但是由于这些变量没有标准的、统一的测量单位,所以它们与对应的观测变量并不是同构的.次序测量尺度上的区间是主观确定的.因此,次序测量尺度只能给出被测量事物特征的相对大小,而不能确切给出一个事物比另一个大多少.

2.6 间隔水平测量

间隔水平测量是比次序水平测量更高的一个水平.它不仅具有以上两个尺度的性质(=,≠,<,>),而且还具有统一的、标准的测量单位.这些测量单位产生具有不变的相等的区间的尺度,从而消除了定量测量中的主观性.使用间隔测量尺度时,通过尺度值之间的加减运算,可以得到两个事物间关于被测量特征的确切差别.间隔尺度得到的测量变量总是与观测变量同构的,但是该测量尺度有**任意**的但不是**绝对的零点**.

例如,温度的**摄氏尺度**就是一种间隔尺度.在这个尺度中,0℃被定义为水的冰点,100℃被定义为水的沸点,而单位(℃)则被定义为二者距离的百分之一.通过加减运算可以确定摄氏测量尺度值之间的确切距离.因此我们可以说40℃的物体比30℃的物体热10℃(40℃−30℃=10℃).

间隔尺度的另一个例子是温度的**华氏尺度**.华氏尺度仍然使用水的冰点和沸点来决定零值和尺度单位.然而,水的结冰点(32℉)与沸点(212℉)之间的距离是180℉(t℉$=\frac{5}{9}(t+32)$℃),而零值(0℉)比水的结冰点低32℉.

2.7 比例水平测量

比例水平测量是测量中的最高水平.它不仅包括名义尺度的性质(=,≠)、次序尺度的性质(<,>)和间隔尺度的性质(+,−),而且还增加了**绝对零点**.也就是说,在比例水平测量中,零值是所测量的物体的特征不能出现或不能观测到的一点.因为在该尺度中,数值表示的是离绝对零点的距离,所以可以通过计算不同测量值间的比例来表示一个测量值是另一个的多少倍.

用以表示温度的**Kelvin尺度**就是一种比例尺度.Kelvin零度是绝对零度,定义为理想气体压强为零(每个气体分子的平均动能为零)时的温度.单位(Kelvin度)所表示的温度长度与摄氏度一样,都是水的冰点与沸点之间距离的1/100,而Kelvin零度是摄氏−273.15℃.在摄氏度和华氏度中计算两个温度间的比例是没有意义的,但在Kelvin尺度中这种计算却是允许的,因此我们可以说300K是150K温度的两倍.

比例测量尺度的其他一些例子如:重量(克)、长度(厘米)、时间(秒)、速度(英里/小时)等许多其他常用的尺度.

例2.2 分别把300K和150K转换为等价的摄氏度,并证明次序水平的摄氏度值间的比例是没有意义的.

解 ☞ Kelvin尺度的绝对零点允许合理的比率,而摄氏零度则不能.为说明,将300K和150K利

用关系式：0℃ =K−273.15 转换为等价的摄氏度.

300K 对应的摄氏度为　300−273.15=26.85

150K 对应的摄氏度为　150−273.15=−123.15

计算二者比得：26.85℃/−123.15℃=−0.218，认为 26.85℃比−123.15℃热−0.218 倍是无意义的.

2.8　连续测量变量和离散测量变量

先举例说明什么是**连续测量变量和离散测量变量**. 考虑某班级中的以下一些测量变量，以厘米测量的学生的身高是一个连续测量变量，而参加各学术报告的学生数则是一个离散测量变量.

学生的身高是一个连续测量变量，因为身高是一个不间断的**连续值**. 从理论上讲，它可以无限细分至任意小数位. 例如，测量某一学生的身高为 165.2 cm，这表明该测量精确到了 1/10 cm. 我们可以通过进一步的手段以得到更加准确的测量值，如 165.195 cm. 依次进行下去，理论上可以用一个具有无限小数位的数值来表示一个无限精确的测量值. 如 165.19534216239867…cm. 对于连续测量变量，理论上（而不是实际上）可以用一个具有无限小数位的数值来表示两个测量值间的差别.

不连续的测量变量就称为离散测量变量，其测量值是有间断的. 最常见的离散测量变量是通过计数得到的. 由于学生数不是连续值，所以参加各学术报告的学生数是一个离散测量变量，如果整个班级的学生数是 52 人，那么参加某学术报告的学生数可能是 52、37 或 25 等，而不可能是 52.1 或 48.639. 对一个离散变量的测量必定是固定的集合中的一个值，而不可能是中间值. 若比例测量得到的是连续测量变量，就称之为**连续比例水平测量**（如学生身高）；反之，若得到的是离散测量变量，则称之为**离散比例水平测量**（如计数）. 计数是比例水平测量，因为它具有名义测量（=，≠）、次序测量（<，>）、间隔测量（+，−）的所有性质，有一个尺度单位（数 1）及绝对零点.

例 2.3　分别按照以下类型给出对图 2−1 中的物体进行测量的一种测量水平：(a)名义尺度，(b)次序尺度，(c)区间尺度，(d)连续比例尺度，(e)离散比例尺度.

解　(a) 一个可能的名义尺度是根据物体的形状将其分为 3 类：三角形、正方形和五角形. 这种尺度将物体分成了一些无序的类别.

(b) 一个可能的次序尺度是按大小把物体分成大、中、小三类，分别用 1、2、3 表示：1=小，2=中，3=大. 这种尺度将物体分成了一些有序的类别.

(c) 一个可能的间隔尺度是物体在 360°圆上的位置. 该尺度以度为单位，但没有绝对零点.

(d) 一个可能的连续比例尺度是物体的最长边的长度，该尺度可以英寸或厘米为单位.

(e) 一个可能的离散比例尺度是图中物体个数的计数.

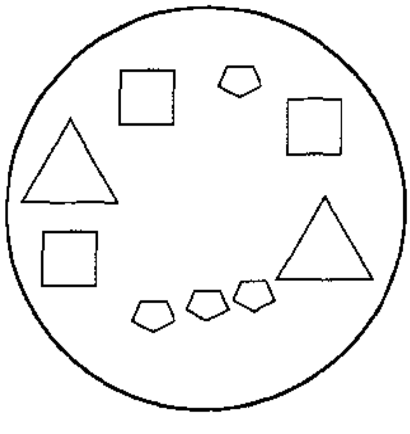


图 2−1

2.9　统计资料的类型

一般而言，**资料**是指为了某项研究而收集的实际信息集合. 在统计中，资料是指测量结果组成的集合，常称为**统计资料**. 根据前面的不同分类，统计资料可相应分为**定性**、**定量统计资料**或者分为**连续**、**离散统计资料**，或按它们的测量水平分，例如**间隔水平资料**等.

对于一组被测量的物体，先用名义水平测量将其分成不同的类别，然后对每一类别中的物体计数. 例如，10 朵花根据名义水平测量分成红色、蓝色和黄色三种类型，通过计数得到红色

的花有3朵,蓝色的花有4朵、黄色的花有3朵.按照这样的计数过程得到的统计资料也称为**分类资料**或**计数资料**.将其称为分类资料是因为物体按照名义尺度被分到不同的类中,将其称为计数资料是因为对分到不同的类中的物体进行了计数.这里的计数方法与2.8节中的离散比例水平测量方法不同,后者得到的计数本身就是测量尺度中的类.

2.10 测量的近似性

在实际中,所有连续统计资料(见2.8节和2.9节)都只是**真实测量值**的一个近似,所以称为**近似测量值**.例如,按照例2.1测量得到某同学的身高为172.7 cm,即172 cm加上172 cm与173 cm之间距离的7/10.然而真实测量值是172 cm与173 cm之间的某个具有无限小数位的确切值.虽然通过增加测量工具的精度,可以得到更多的小数位数从而更接近真实测量值.但是,即使是最精确的测量工具其测量的精度也是有限的,所以说所有的连续统计资料都是近似测量值.

对于每个近似测量值,其最后一位数字(常称为**怀疑数字**)受测量工具的精度的限制.例如对于身高测量值172.7 cm,前三位数字是由米尺直接得到的,是完全确定的,但是其最后一位数字是不确定的估计值:172 cm与173 cm之间距离的7/10.这个估计值事实上当作是一个区间,称为**蕴含范围**.这个范围是以估计值为中点,上下扩展最小尺度单位的1/2得到的.在这个例子中,最小单位是0.1 cm,所以蕴含范围是从172.65 cm到172.75 cm.这种形式是表示蕴含范围的一般形式,在本书中就采用这种表示形式.需要指出的是,有的统计学家为了避免相邻区间的边界重叠而将蕴含范围表示成172.65000…cm到712.74999…cm.

例2.4 指出测量值1.2965 mg的蕴含范围.

解 ☞ 确定近似测量值的蕴含范围的方法为:

(1)根据最小测量单位建立一个含三个数字的测量尺度:测量值为中点、上下各扩展一个最小单位;

(2)分别计算这三个数中前两个数和后两个数的中点,即得蕴含范围的下界和上界.

所以

三个数字的测量尺度	中点	蕴含范围
1.2964		
	← 1.29645	
1.2965		1.29645 mg ~ 1.29655 mg
	← 1.29655	
1.2966		

离散比例水平测量(见2.8节)通常产生**准确**的测量结果.如一个鸟巢中蛋的个数可能恰好为7,而不可能是$6\frac{1}{2}$和$7\frac{1}{4}$.但是大量物体的计数往往是估计的.例如土壤学家在估计一立方米土壤中有机体的个数时,他得到的结果可能是36,000(精确到1,000),经济学家估计美国6月份开始建造的新住宅为230,000间(精确到10,000).

2.11 有效数字

测量中的**有效数字**是在测量尺度中实际能够得到的所有数字.非有效数字是那些在测量中只用来表示小数点的位置的零.对于大于或等于1的数,所有的非零数字是有效的,有效数字之间的零也是有效的,如果这个数包含小数点,那么最后一位非零数字的后边所有的零也都是有效数字.对于0和1之间的数字,第一个非零数字左边的所有零都是非有效数字,其他数字都是有效数字.

例2.5 指出下列近似测量有几位有效数字:(a) 1.12 mg,(b) 1.02 g,(c) 920.02080 mi,(d) 0.0900 mg.

解 ☞ (a) 3,(b) 3,(c) 8,(d) 3

例 2.6　指出下列测量值的蕴含范围和有效数字：(a) 102 匹马，(b) 10,100 只昆虫(精确到 100)

解 ☞　(a) 102 匹马是一个准确数值，所以没有蕴含范围. 它有三位有效数字.

(b) 10,100 只昆虫(精确到 100)是一个估计的离散比例测量值，其蕴含范围为 10,050 只昆虫到 10,150 只昆虫. 它有三位有效数字.

2.12　科学记数法和数量级

科学记数法将一个数表示为两个因子的乘积. 第一个因子是绝对值大于等于 1 且小于 10 的数，第二个因子是 10 的整数次幂. 例如，100.0 用科学记数法表示就是 1.000×10^2，0.36 用科学记数法表示就是 3.6×10^{-1}. 当一个数用科学记数法表示时，10 的指数反映了该数的大小，故称之为此数的**数量级**. 若两个数的 10 的指数相同，那么它们具有相同的数量级. 例如，1.0×10^3、4.213×10^3 和 9.237456×10^3 的数量级相同.

例 2.7　对于下面用科学记数法表示的近似测量值，指出其有效数字和蕴含范围：4.2×10^3 m.

解 ☞　习惯上，科学记数法中所有的有效数字均写在第一个因子里，并且其中没有非有效数字. 因此，该测量值的有效数字是 4 和 2.

在确定用科学记数法表示的数值的蕴含范围时，对于第一个因子按照例 2.4 给出的方法得到两个数，然后再分别乘以第二个因子，即得到蕴含范围的两个端点. 所以该数的蕴含范围为 4.15×10^3 到 4.25×10^3.

2.13　测量的系统误差和随机误差

在 2.10 节中，我们指出所有连续统计资料都只是真实测量值的一个近似，所以称为近似测量值. 原因之一是即使对最精确的测量工具也有精确度的限制. 另外一个原因是测量中出现的误差. 这些误差可能是**系统误差**，也可能是**随机误差**.

测量的系统误差也称为**偏差**，它是由测量过程中的缺陷造成的，测量过程中总是产生向一个方向偏离的数据，导致测量结果要么总是太大、要么总是太小. 如例 2.1 中，由于某种原因，上方米尺的底端刻度线比下方米尺的顶端刻度线低 1 厘米，将导致测量结果比实际值总是大 1 厘米. 这种误差通常很难察觉，但是可以通过将测量值与一些经验值或理论标准值相比较来发现这种误差.

测量中出现的另一种误差是**随机误差**(或**机会误差**)，这类误差是由于测量过程的随机变异性导致的，测量结果有时偏大，有时偏小. 例如，对同一个学生的身高测量十次，学生的站立姿势或测量者的观测角度的变化，都有可能产生不一致的偏差，从而产生随机误差.

由于精确度的限制和误差的存在，所以每一个近似测量值可以用如下等式表示：

近似测量值＝受精确度限制的真实测量值＋系统误差＋随机误差

如果离散比例水平测量(见 2.8 节)是用于对确定数量物体的计数，那么测量结果不可能包括系统误差和随机误差；但是，若是对物体数量进行估计(见 2.10 节)，则测量结果就会包括系统误差和随机误差.

2.14　统计中的准确度和精密度

在统计中，测量值的**准确度**是指测量值与真实值之间的一致程度. 由上节可知，准确度是由测量工具的精度和测量误差(尤其是系统误差)所决定的.

精密度与准确度不同，它是指对同一事物进行重复测量所得测量结果的一致程度，也简称为**精度**. 在每次测量中，系统误差一般是保持不变的，所以精密度，或重复测量的变异性，主要取决于随机误差的存在及其大小.

通过以上定义，我们可以看出测量结果可能是：既准确又精密(靠近真实值并且重复测量

的变异性小)、既不准确又不精密(远离真实值并且重复测量变异性大)、准确但不精密、精密但不准确.

近似测量中的有效数字表明测量的准确度.例如,测量得到某一物体的重量是 4.32 mg,这表明该测量值只能准确到 4.3 mg,并且真实值在 4.315 mg 到 4.325 mg 的蕴含范围中.由于涉及准确度,所以蕴含范围又通常称为准确度的**蕴含范围**.

2.15 自然科学中的准确度和精密度

上一节中关于准确度和精密度的统计定义来自第十四章介绍的统计估计理论.在自然科学(化学、物理学等)中用到的有关测量的准确度和精密度的定义与其类似但稍有不同.

在自然科学中,测量的准确度水平是用测量值的有效数字的个数表示的,而测量的精密度水平则是由最小测量单位的大小确定的.例如,4.32 mg、43.2 mg 和 432 mg 都有 3 个有效数字,所以它们都是在相同的准确度水平下测量的;而 423.3 mg、43.2 mg 和 4.3 mg 都精确到 0.1 mg,所以它们都是在相同的精密度水平下测量的.

例 2.8 今有 3 个用科学记数法表示的测量值:2.531×10^2 cm、2.531×10 cm、2.5316×10^2 cm 和 2.53167×10^3 cm.找出哪些测量值具有相同的(a)准确度,(b)精密度,(c)数量级.

解 ☞ (a) 2.531×10^2 cm 与 2.531×10 cm 具有相同的准确度水平;

(b) 2.531×10 cm、2.5316×10^2 cm 和 2.53167×10^3 cm 具有相同的精密度水平,都精确到 0.01 cm;

(c) 2.531×10^2 cm 和 2.5316×10^2 cm 的数量级相同.

近似测量值进行加减运算时,所得结果应进行舍入运算(见 1.4 节),使其与最低精密度(自然科学中的定义)的测量值具有相同的精密度水平;进行乘除运算时,所得结果也应进行舍入运算,使其与最低准确度(自然科学中的定义)的测量值具有相同的准确度水平.

例 2.9 完成指定运算并将结果舍入到正确的位数:(a) 7.123 kg+8.9 kg,(b) 72 kg×0.01 kg.

解 ☞ (a) 7.123 kg+8.9 kg=16.023 kg,并舍入到 16.0 kg

(b) 72 kg×0.01 kg=0.72 kg^2,并舍入到 0.7 kg^2

2.16 单位换算

米制又称为**国际单位制**,其主要单位有:米(m)、克(g)和秒(sec).这些单位的标准倍数和约数都有特定的名称及缩写.

例 2.10 写出下列各量的名称和缩写:(a) 10^{-2} m,(b) 10^{-3} sec,(c) 10^{-6} g,(d) 10^{-9} m,(e) 10^3 g,(f) 10^6 sec,(g) 10^9 m.

解 ☞ (a) 10^{-2} m 的名称是厘米,缩写为 cm

(b) 10^{-3} sec 的名称是毫秒,缩写为 msec

(c) 10^{-6} g 的名称是微克,缩写为 μg

(d) 10^{-9} m 的名称是纳米,缩写为 nm

(e) 10^3 g 的名称是千克,缩写为 kg

(f) 10^6 sec 的名称是兆秒,缩写为 Msec

(g) 10^9 m 的名称是千兆米,缩写为 Gm

米制单位与**英制单位**之间的换算是通过换算因子(见表 2-1)完成的.

表 2-1

换算因子		换算因子	
英制→米制	米制→英制	英制→米制	米制→英制
1 inch=2.540 cm(厘米)	1 cm=0.3937 in(英寸)	1 lb=0.4536 kg(千克)	1 kg=2.205 lb(磅)
1 inch=2.540×10^{-5} km(千米)	1 km=3.937×10^{4} in(英寸)	1 lb=453.6 g(克)	1 g=2.205×10^{-3} lb(磅)
1 ft=0.3048 m(米) 1 mi=1.609 km(千米)	1 m=3.281 ft(英尺) 1 km=0.6215 mi(英里)	1 gal=3.785 liters(l)(升)	1 l=0.2642 gal(加仑)

例 2.11　将 100yd(码)换算成以厘米为单位的量.

解 ☞　$$100\text{yd}\times\left(3\,\frac{\text{ft}}{\text{yd}}\right)\times\left(12\,\frac{\text{in}}{\text{ft}}\right)\times\left(2.540\,\frac{\text{cm}}{\text{in}}\right)=300\text{ ft}\times\left(12\,\frac{\text{in}}{\text{ft}}\right)\times\left(2.540\,\frac{\text{cm}}{\text{in}}\right)$$
$$=3{,}600\text{ in}\times\left(2.540\,\frac{\text{cm}}{\text{in}}\right)$$
$$=9{,}144\text{ cm},\text{或 }9.144\times10^{3}\text{ cm}$$

在计算中唯一的近似数是换算因子,它有四位有效数字,所以结果不需进行舍入.

习 题 解 答

测量水平和测量单位

2.1　为什么在自然科学中将测量限制为间隔水平测量和比例水平测量?

解 ☞　在化学、物理和其他自然科学中,所有的测量都必须包括数值和单位,例如 20 英尺、15 磅,即采用比例测量或间隔水平测量.然而在大多数社会科学中,例如心理学、社会学、经济学等,测量通常都取为名义和次序这两种较低水平、没有单位的测量.所以本书中,我们采用了包括了名义水平和次序水平的测量的一般定义.

2.2　用例 2.3(a)中提出的名义尺度对图 2-1 中的物体进行计数.

解 ☞　对于名义尺度中的三个类别,每类的个数分别是:2 个三角形、3 个正方形和 4 个五边形.

2.3　请判断下列各种测量的水平,并说明原因:(a)用数字 1~5 表示美国人对移民的态度,其中 1=不欢迎,…,5=非常欢迎,(b)某商店 40 名职员的性别,(c)每天来吃食的鸟的类型,(d)同年出生的 50 个学生的出生日期,(e)女运动员完成百米冲刺的时间,(f)用℃表示的人体温度.

解 ☞　(a) 次序尺度.因为这 5 个数字表示的态度是有序的类,但是各类之间的划分是主观确定的.

(b) 名义尺度.因为性别(男性、女性)是无序的两类别尺度.

(c) 名义尺度.因为鸟的类型是无序的多类别尺度.

(d) 间隔尺度.因为日期有测量单位(天),但是无绝对零点.

(e) 连续比例尺度.因为时间长度有测量单位(秒),且有绝对零点(跑步开始的时间).

(f) 间隔尺度.因为人体温度有标准测量单位(℃),但是无绝对零点.

2.4　请判断下列各种测量水平:(a)某地区范围每加仑汽油的价格,(b)维他命的类型(例如 VE),(c)每个足球队成员的球衣号,(d)用数字 1~4 测量苹果的甜度,其中 1=不甜,…,4=非常甜,(e)一堆苹果中每个苹果的糖含量(以毫克为单位).

解 ☞　(a) 离散比例尺度.

(b) 名义尺度.

(c) 名义尺度.

(d) 次序尺度.

(e) 连续比例尺度.

测量的近似性

2.5　写出近似测量值 0.19032 cm 的蕴含范围.

解☞ 用例 2.4 中的方法得到蕴含范围的端点为 0.190315 cm 和 0.190325 cm.

2.6 写出近似测量值 700.3 kg 的蕴含范围.

解☞ 蕴含范围的端点为 700.25 kg 和 700.35 kg.

2.7 写出下列近似测量值的蕴含范围:(a) 1,000,000.0 g,(b) 0.0001 m.

解☞ (a) 999,999.95 g 到 1,000,000.05 g

(b) 0.00005 m 到 0.00015 m.

有效数字

2.8 指出下列近似测量有几个有效数字:(a) 1.20 kg,(b) 1.00000 cm,(c) 0.0056 mg,(d) 0.04003 mg.

解☞ (a) 3,(b) 6,(c) 2,(d) 4

2.9 指出下列测量值的蕴含范围和有效数字个数:(a) $10\dot{0}$,000 棵树,(b) 100,001 棵树,(c) 100,000 棵树.

解☞ (a) 对于无小数点的数字,通常用最后一个有效数字上加点的方法表示有效数字.所以,$10\dot{0}$,000棵树有三个有效数字,蕴含范围为:99,500 棵树到 100,500 棵树.

(b) 100,001 棵树是一个准确值,无蕴含范围,有六个有效数字.

(c) 另一种表示无小数点数字的有效零的方法是将非有效零用更小字体表示.所以,100,000 棵树有三个有效数字,蕴含范围为 99,500 棵树到 100,500 棵树.

2.10 指出下列测量值的蕴含范围和有效数字个数:(a)10,000 个工人,(b)103,000.0 英里.

解☞ (a)该测量值没有给出足够的信息以确定蕴含范围和有效数字个数.

(b)蕴含范围为:102,999.95 英里到 103,000.05 英里,有七个有效数字.

科学记数法和数量级

2.11 对下面用科学记数法表示的近似测量值,指出哪些是有效数字,蕴含范围是什么:8.7961×10^{-2}米?

解☞ 该测量值的有效数字是 8,7,9,6 和 1.蕴含范围为:8.79605×10^{-2}米到 8.79615×10^{-2}米.

2.12 对下列这些用科学记数法表示的近似测量值,指出哪些是有效数字,蕴含范围是什么:(a) 9.99×10^{-6} kg,(b) 2.0×10^{6} kg?

解☞ (a)该测量值的有效数字是 9,9,9,蕴含范围为:9.985×10^{-6} kg 到 9.995×10^{-6} kg.

(b) 该测量值的有效数字是 2,0.蕴含范围为:1.95×10^{6} kg 到 2.05×10^{6} kg.

2.13 用科学记数法表示以**小数记号**写出的数:0.000000060 mm.

解☞ 要将小于 1 的小数用科学记数法表示,向右移动小数点直到第一位不为零的数字的右边.对于 0.000000060 mm,小数点向右移了 8 位.因此第一个因子为 6.0,第二个因子是 10^{-8}.所以,用科学记数法表示此数为:6.0×10^{-8}mm.

2.14 用科学记数法表示以**小数记号**写成的数:4,000,000,000.0 kg.

解☞ 要将 10 或更大的数用科学记数法表示,向左移动小数点直到第一位不为零的数字的左边.对于 4,000,000,000.0 kg,小数点向左移动 9 位.因此第一个因子是 4.所以,用科学记数法表示此数为 4.0000000000×10^{9} kg.

2.15 用小数表示以科学记数法写出的数 4.92×10^{-7} m.

解☞ 用小数表示此数为 0.000000492 m.

2.16 用小数表示以科学记数法写出的数 6.0×10^{7} kg.

解☞ 用小数表示此数为 60,000,000 kg.

自然科学中的准确度和精密度

2.17 完成指定运算并将结果舍入到正确的位数：(a)9.99623 kg－8.12 kg，(b)9.99 kg÷8 kg，(c) 4.23 kg×100.0039 kg.

解 (a) 9.99623 kg－8.12 kg＝1.87623 kg，应舍入到 1.88 kg，

(b) 9.99 kg÷8 kg＝1.24875，应舍入到 1，

(c) 4.23 kg×100.0039 kg＝423.016497 kg^2，应舍入到 423 kg^2

2.18 完成指定运算并将结果舍入到正确的位数：(a) 91.26 g × 1.1 g，(b) 452.1 g － 21.239 g.

解 (a) 91.26 g×1.1 g＝100.386 g，应舍入到 100 g，

(b) 452.1 g－21.239 g＝430.861 g，应舍入到 430.9 g

2.19 完成指定运算并将结果舍入到正确的位数：(4 mm－2.92 mm)×8.397 mm.

解 首先 4 mm－2.92 mm＝6.92 mm，应舍入到 7 mm，

然后 7 mm×8.397 mm＝58.779 mm^2，应保留一个有效数字，舍入到 60 mm^2.

2.20 完成指定运算并将结果舍入到正确的位数：72.916 mm×4.21 mm－6 mm^2.

解 首先 72.916 mm×4.21 mm＝306.97636 mm^2，应舍入到 307 mm^2，

然后 307 mm^2－6mm^2＝301 mm^2，不需进行舍入.

2.21 完成指定运算并将结果舍入到正确的位数：$(3.926\times10^2\ kg)\times(4.29\times10^3\ kg)$.

解

$$
\begin{aligned}
(3.926\times10^2\ kg)\times(4.29\times10^3\ kg) &= (3.926\times4.29)(10^2\times10^3)(kg\times kg)\\
&= 16.84254\times10^5\ kg^2\\
&= 1.684254\times10^6\ kg^2 = 1.68\times10^6\ kg^2
\end{aligned}
$$

最后应舍入成 $1.68\times10^6\ kg^2$.

2.22 完成指定运算并将结果舍入到正确的位数：$(2.9\times10^{-3}\ ft)+(4.26\times10^{-4}\ ft)$.

解

$$
\begin{aligned}
(2.9\times10^{-3}\ ft)+(4.26\times10^{-4}\ ft) &= 0.0029\ ft+0.000426\ ft\\
&= 0.003326\ ft, 0.0033\ ft, \text{或 } 3.3\times10^{-3}\ ft.
\end{aligned}
$$

最后应舍入成 0.0033 ft 或 3.3×10^{-3} ft.

单位换算

2.23 某住宅的一侧边长为 9 米，分别以(a)厘米，(b)毫米为单位来表示这个距离.

解 (a) $9m\times\left(\frac{100\ cm}{1\ m}\right)=900\ cm$

(b) $9m\times\left(\frac{1{,}000\ mm}{1\ m}\right)=9{,}000\ mm$

2.24 用适当的换算因子将 350 千米换算成以英尺为单位的量.

解

$$350\ km\times\left(1{,}000\ \frac{m}{km}\right)\times\left(3.281\ \frac{ft}{m}\right)=(3.50000\times10^5\ m)\left(3.281\ \frac{ft}{m}\right)$$

应舍入成 1.148×10^6 ft.

2.25 用适当的换算因子将 12.9 千克换算成以磅为单位的量.

解

$$12.9\ kg\times\left(2.205\ \frac{lb}{kg}\right)=28.4445\ lb$$

应舍入成 28.44 lb.

2.26 用适当的换算因子将 9,920 磅换算成以千克为单位的量.

解

$$(9.920\times10^3\ lb)\left(4.536\times10^{-1}\frac{kg}{lb}\right)=4.499712\times10^3\ kg$$

应舍入成 4.500×10^3 kg.

2.27 用适当的换算因子将 20 加仑换算成以升为单位的量.

解

$$20\ gal\times\left(3.785\ \frac{liter}{gal}\right)=75.70\ liters$$

补 充 习 题

测量水平和测量单位

2.28 指出下列各种测量的水平:(a)以毫米为单位测量蜗牛壳的直径,(b)用数字 1~6 对学生的散文进行评分,其中 1－非常好,…,6－差,(c)某制造商的客车的年销售量.

答案 (a)连续比例,(b)次序,(c)离散比例

2.29 指出下列各种测量的水平:(a)每分钟录入的字数,(b)－273.15℃,(c)小孩用右手或左手的习惯.

答案 (a)离散比例,(b)间隔,(c)名义

2.30 指出下列各种测量的水平:(a)用四种颜色表示铁的温度(灰色＝冷＝1,黄色＝暖＝2,红色＝热＝3,白色＝炽热＝4),(b)某产品每月的销售额(以美元为单位),(c)失业工人数.

答案 (a)次序,(b)离散比例,(c)离散比例

测量的近似性

2.31 指出下列近似测量值的蕴含范围:(a) 4,926.22 cm,(b)0.1920 sec,(c)41.00001 in.

答案 (a) 4,926.215 cm 到 4,926.225 cm

(b) 0.19195 sec 到 0.19205 sec

(c) 41.000005 in 到 41.000015 in

有效数字

2.32 指出下列近似测量值有多少个有效数字:(a) 4,930,200.0 km,(b) 0.8001 mg?

答案 (a)8,(b)4

2.33 指出下列近似测量值有多少个有效数字:(a) 0.0006000 mg,(b) 0.000047 mg?

答案 (a)4,(b)2

2.34 指出下面测量值的蕴含范围和有效数字个数:52,000 只飞蛾?

答案 蕴含范围为:51,500 只飞蛾到 52,500 只飞蛾,有两个有效数字.

2.35 指出下面测量值的蕴含范围和有效数字个数:94,000 英里(精确到 100)?

答案 蕴含范围为:93,950 英里到 94,050 英里,有三个有效数字.

科学记数法和数量级

2.36 对下面用科学记数法表示的测量值,指出哪些是有效数字,蕴含范围是什么:1.0000×10^{-1} kg?

答案 该测量值的有效数字是 1,0,0,0 和 0,蕴含范围为:9.9995×10^{-2} kg 到 1.00005×10^{-1} kg.

2.37 用科学记数法表示 80,888 m.

答案 8.0888×10^4 m.

2.38 用科学记数法表示 0.90009 lb.

答案 9.0009×10^{-1} lb.

2.39 把下列用科学记数法表示的测量值用小数表示,用小数表示的则用科学记数法表示:(a) 2.000×10^{-2} mm,(b) 1,001.00 kg.

答案 (a) 0.02000 mm,(b) 1.00100×10^3 kg.

2.40 用小数表示:(a) 8.11×10^5 sec,(b) 5.1×10^{-9} in.

答案 (a)811,000 sec,(b) 0.0000000051 in.

自然科学中的准确度和精密度

2.41 完成指定运算并将结果舍入到正确的位数:(a) 0.39247 kg÷0.0000007 kg＋0.21 kg,(b) 2.1 kg÷0.000056.

答案 (a) 0.6024707 kg,应舍入到 0.60 kg,

(b) 37,500.0 kg,应舍入到 38,000 kg.

2.42 完成指定运算并将结果舍入到正确的位数:(a) 1.26 g÷312.92,(b) 892 g＋2.263 g.

答案 (a) 0.00402659 g,应舍入到 0.00403 g,

(b) 894.263 g,应舍入到 894 g.

2.43 完成指定运算并将结果舍入到正确的位数:(6.1 in×2.936 in)÷18.23914 in^2.

答案 18 in^2÷18.23914 in^2=0.99.

2.44 完成指定运算并将结果舍入到正确的位数:3.2937 in^2−22.3 in^2+8.421 in×39.213 in.

答案 −19.0 in^2+330.2 in^2=311.2 in^2.

2.45 完成指定运算并将结果舍入到正确的位数:(1.926×105 kg)÷(9.1×10^4).

答案 2.11648 kg,应舍入成 2.1 kg.

2.46 完成指定运算并将结果舍入到正确的位数:(5.27×10^3 ft)−(8.838×10^2 ft).

答案 4,386.2 ft,应舍入成 4,390 ft 或 4.39×10^3 ft.

单位换算

2.47 对于习题 2.23 中的住宅,把 9 米分别以(a)微米,(b)千米为单位来表示.

答案 (a) 9,000,000μm,(b) 0.009 km

2.48 用适当的换算因子将 111 英里换算成以千米为单位的量.

答案 178.599 km,应舍入成 178.6 km.

2.49 用适当的换算因子将 103 克换算成以磅为单位的量.

答案 2.27115×10^{-1} lb,应舍入成 2.271×10^{-1} lb.

2.50 用适当的换算因子将 1,350 升换算成以加仑为单位的量.

答案 356.670 gal,应舍入成 356.7 gal.

第三章　总体、样本和统计量

3.1　自然总体和测量总体

总体有很多非统计的含义.在生物学中,是指居住在同一地理区域并能相互交配的同一物种的个体构成的整体.在社会学中,是指居住在同一国家、地区或社区的所有的人.在物理学中,是指在同一能量级上的所有的粒子.

在统计学中,总体具有特殊的含义,它与对测量资料进行分析这一统计基本任务密切相关.如前所述,每一个特定的测量都有一个相应的操作定义(见 2.2 节),这一定义必须包括对被测量事物的确切描述:如年龄、时间、温度、地点和其他进一步的特征.符合这一描述的所有事物(包括过去、现在、将来)的集合(见 1.17 节)称为该测量的**自然总体**.自然总体中每个元素的测量值集合称为**测量总体**.由于资料分析的统计方法主要针对测量总体,所以**从本章开始若没作特别说明,均指测量总体**.

为了理解这些概念,我们考虑一个具体的例子.植物遗传学家要研究一种新型玉米,需要在成熟期测量玉米苗的高度.在这个测量的操作定义中,除了说明测量的具体步骤外,还需对要测量的玉米苗进行准确地描述:如何选种、苗的生长条件(包括土壤环境、种植方法、施肥及灌溉方案等)以及测量苗高的时间为种植后 4 个月.所有在这些条件下种植的玉米苗就构成了该测量的自然总体.对这个自然总体中每个元素进行测量所得的高度值的集合就是测量总体.

3.2　有限总体、无限总体和假设总体

总体中的每一项称为**元素**.无论是自然总体还是测量总体,如果其元素个数具有上限,则称该总体为**有限总体**;否则称为**无限总体**.

例如,参加某届奥运会百米决赛的 8 个人就构成一个有限自然总体,与该有限自然总体相联系的有限测量总体是每个人的决赛成绩.满足 3.1 节描述的种植 4 个月的玉米苗(过去、现在、将来)就是一个无限自然总体,而它们的高度就是相应的无限测量总体.一般情况下,无限总体并不实际存在,而只是假设为无限总体,所以也称为**假设总体**(或**虚拟总体**,或**概念总体**).

例 3.1　掷骰子并数出最上面的点数,请问:(a) 被测量物体是什么? (b) 所采用的测量尺度是什么? (c) 这一测量的自然总体是什么? (d) 相应的测量总体是什么? (e) 自然总体和测量总体是有限的还是无限的?

解 ☞　(a) 被测量物体是骰子落下后朝上的一面.

(b) 所采用的测量尺度是关于点数的离散比例尺度(见 2.8 节).

(c) 这一测量的自然总体是包括所有骰子落下后朝上的一面的假设总体.

(c) 相应的测量总体是上述自然总体的所有点数.

(e) 自然总体和测量总体均是无限的.

3.3　样本

同总体概念一样,统计学中的**样本**也包括**自然样本**和**测量样本**.自然样本是在自然总体中选取的部分元素组成的子集,而对自然样本中的所有元素的测量构成一个测量样本.由于后面各章中的总体指的是测量总体,所以**样本指的就是测量样本**.

通常,总体和样本的区别与所研究问题的范围很有关系.例如,一个大城市中的失业劳动力,如果所研究的问题是这个特定城市的经济状况,则它为自然总体;如果所研究的是一个地区或国家的经济状况,则它为自然样本.

例 3.2　请问下列样本是自然样本还是测量样本：(a) 500 个家庭中的 15 个家庭的年收入，(b) 一个肉加工厂一周生产的香肠中的 10 节香肠的脂肪含量，(c) 一个汽车生产厂的每天生产汽车中的 5 辆，(d) 1985 年以来，美国的囚犯中的 10 个.

解 ☞　(a) 测量样本

(b) 测量样本

(c) 自然样本

(d) 自然样本

3.4　参数与统计量

测量值集合具有自身可以被测量和描述的特征，例如：最小值与最大值之间的距离，数据是均匀地散布在最大值与最小值之间的范围内还是聚集在一个或 n 个位置上，集合中最典型、最具代表性的值，等等. 任何一个由所有测量总体计算得到的描述这一总体的特征的数值称为**参数**(或**总体参数**). 同样的，由测量样本得到的描述性的值称为**统计量**(或**样本统计量**，或**统计度量**).

最常用的描述性度量是**算术平均数**，它描述了统计资料集合的平均值. 在第六章中，我们将详细地讨论这一特征数以及**集中趋势**、**平均值**和**位置**的度量等. 算术平均数是由集合中所有测量值的和除以测量值的个数得到的. 如果该集合是测量样本，算术平均数是按下式计算得到的一个统计量

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n} \tag{3.1}$$

其中，$\bar{x}$ 表示样本均值，n 表示**样本容量**(测量样本中元素的个数)，$\sum_{i=1}^{n}x_i$ 是常用的求和记号(见 1.12 节).

如果该集合包括所有测量总体，则算术平均数是按下式计算得到的一个参数.

$$\mu=\frac{\sum_{i=1}^{N}x_i}{N} \tag{3.2}$$

其中，μ(希腊字母)代表总体均值，N 代表**总体容量**(测量总体中元素的个数).

本书将给出很多这种对应的参数和统计量，一般参数用希腊字母表示，统计量用罗马字母表示. 而且，测量总体的参数值是唯一确定的，而统计量的值是随测量样本的不同而改变的.

例 3.3　计算下列算术平均数：(a) 测量样本 $x_1=7$、$x_2=5$，$x_3=6$，$x_4=6$，(b) 测量总体 $x_1=0.2$、$x_2=0.6$，$x_3=0.4$.

解 ☞　(a) $\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{\sum_{i=1}^{4}x_i}{4}=\frac{7+5+6+6}{4}=6$

(b) $\mu=\frac{\sum_{i=1}^{N}x_i}{N}=\frac{\sum_{i=1}^{3}x_i}{3}=\frac{0.2+0.6+0.4}{3}=0.4$

3.5　统计科学

统计一词有很多含义. 它可以表示样本资料的数值的描述性度量(见 3.4 节)，也可以表示资料的收集，或者是关于测量资料收集、整理、分析和推断的科学. 最后一种含义，即统计学，是本书的主题.

在历史记载中，人们很早就开始收集资料. 统计(statistics)一词就是来源于拉丁语中的国家(status)，最早指的是一个国家政府要求的来自各个地区的资料，例如，税收、农作物的种植等等. 现在，统计的含义已经扩展到包括任何资料的收集，例如，劳动力统计，体育资料统计等.

统计科学不只是纯粹的资料积累.原始收集到的统计资料常常是大量的,人们很难直接使用它们进行研究.统计科学为数据资料的使用提供了理论和方法支持.正如1.1节所提到的,统计学分为两部分:**描述统计学**和**推断统计学**.

上册讨论描述统计学的基本原理:物体的测量(见第二章)、将统计资料整理成表格(见第四章)和图形(见第五章)、计算描述统计量和参数(见第六、七章).上册的后半部分及下册则介绍推断统计学的基本原理.推断统计学为由样本特征**推断**总体特征提供了方法和逻辑.除了这种从样本到总体的推断方法,它也包括概率论(见第八章到第十二章)和抽样理论(见第十三章),用来确定推断的优劣,即这种推断靠近真实值的程度.

3.6　估计问题和假设检验问题

推断统计学的理论基础是**统计决策理论**(简称**决策理论**).该理论主要包括:**估计理论**和**假设检验理论**.

估计理论处理的是估计问题.在这类问题中,测量总体的未知参数是用相应测量样本的统计量来估计的,例如:用样本均值 $\bar{x}$ 估计总体均值 μ.除此之外,估计问题的完整解答还应包括估计量的优劣程度,即它的确定程度,通常用包含估计值的一个区间来表示,这个区间称为**置信区间**.估计理论将在第十四章和第十五章中介绍.

假设检验理论解决的是**统计假设**应接受还是拒绝的问题.统计假设是指关于一个或多个测量总体的未知性质,特别是关于它们的参数或它们是如何散布的假设或猜测.考虑参数的假设,例如,总体均值等于某常数 $a(\mu=a)$,则假设检验使用样本信息以确定假设的真实程度.单个测量样本的假设检验理论将在第十六章讨论.

图3-1给出了单个总体的某个样本的估计和假设检验的步骤.图中测量总体的均值 μ 是未知的.为了估计并进行假设检验,首先从自然总体中抽取一个样本,然后对这一自然样本中的每个元素进行某种测量,得到一个测量样本,由这个测量样本计算出样本均值和其他描述统计量,最后运用推断统计学确定 μ 的置信区间,以及统计假设 $\mu=a$ 的真实程度.

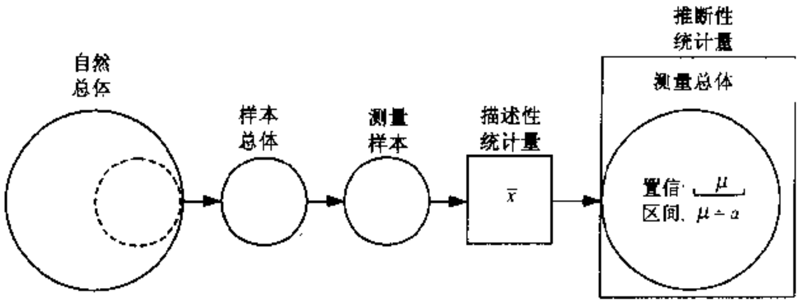


图3-1

3.7　统计假设和研究假说

统计假设(见3.6节)和**研究假说**(也称为**科学假说**,或**工作假说**)在某些方面是相似的,但是在其他方面却是截然不同的.它们都是有关总体的假设.研究假说是有关自然总体的假设,而统计假设则是有关测量总体的假设(例如,$\mu=a$).统计假设是对一些测量集的特征的抽象假设,研究假说处理的则是实际的因果关系问题(见1.19节),例如,根据对肺癌患者的研究,医学研究者提出的因果研究假说是:吸烟导致肺癌.类似的,儿童社会学家提出这样的假说:青少年过度看电视会导致暴力倾向.

统计假设和研究假说都是通过**演绎推理**(见习题3.6)导出结果进行检验的:如果假设成立,那么会出现什么结果.检验统计假设时所用的数学的演绎推理是假设检验理论的一部分,我们将在第十六章介绍.然而,检验研究假说时所用的演绎推理并没有合适的理论支持、研究

者必须自己根据其假设来完成逻辑推理.下例是一个较明显的演绎推理:如果吸烟导致肺癌成立,那么在吸烟人群中肺癌患者要比不吸烟人群中的肺癌患者多.

统计假设是用来对测量总体特征作出决策的抽象的数学工具.统计假设一旦成立,这些特征将被用于对研究假说的决策.

3.8　探索性研究和假设检验研究

任何领域认识的发展都是一个研究假说的形成和检验的过程.**探索性研究**(也称为**描述性研究**)是假说的形成阶段,在这个阶段,人们根据观察、测量和数据分析并通过归纳推理(见习题3.6)来形成研究假说.相反的,**假设检验研究**,如它们的名称所表示的,则是通过检验它们的预测来检验已形成的研究假说.

3.9　探索性试验

在例1.31中,我们指出在试验中是通过自变量的值的改变(控制)来研究其对因变量变化的影响.仍考虑鲑鱼的生长试验,我们饲养两组鲑鱼,一组水温控制在20℃,另一组水温则控制在24℃,其他条件都相同.这个试验中水温是自变量,而鲑鱼的生长情况(用200天后的体重来测量)是因变量.

我们将自变量的取值称为**变量的水平**,也称为**处理**.在鲑鱼的生长试验中,变量的水平是20℃和24℃.我们可以说将自变量(水温)的两种水平应用于鲑鱼,观察它们对因变量(体重)的影响,或者说对鲑鱼应用了两种水温处理.

探索性试验是探索性研究的一种形式.这种试验的目标是通过试验结果形成研究假说.例如,通过鲑鱼的生长试验,生物学家能形成有关水温对鲑鱼的生长作用的假设.

图3-2给出了对鲑鱼的生长试验数据进行分析的步骤.研究者感兴趣的是,两种自变量水平(μ_{20}和μ_{24})下鲑鱼在200天后的平均体重,以及这两个平均体重是否显著不同.

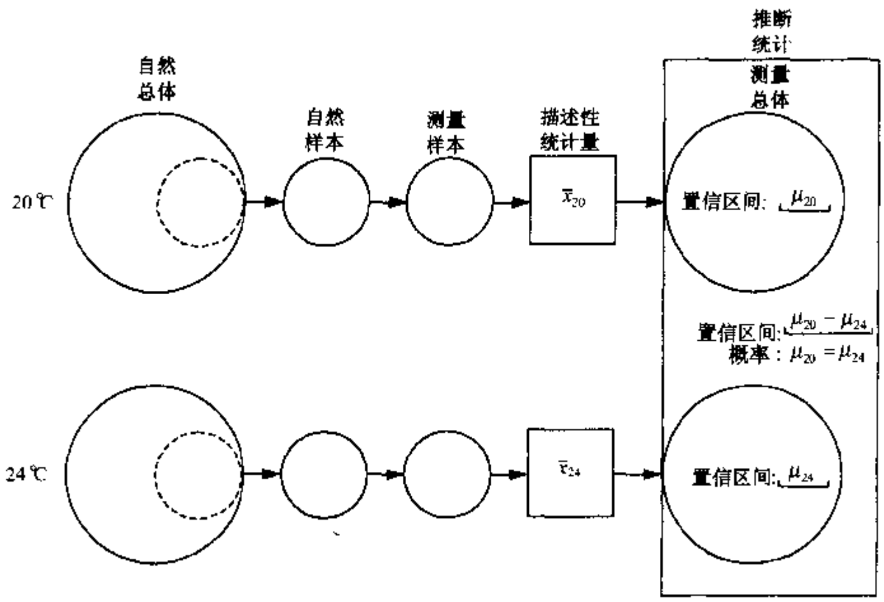


图3-2

这两组鲑鱼是来自无限自然总体的自然样本.对两组鲑鱼中的每一条鲑鱼进行称重就得到两个测量样本,根据描述统计学,由这两个测量样本可以得到两个样本均值,$\bar{x}_{20}$和$\bar{x}_{24}$以及其他样本信息.根据推断统计学,可以得到每个测量总体的均值估计(μ_{20}和μ_{24}的置信区间),以及均值差的估计($\mu_{20}-\mu_{24}$的置信区间),并且对统计假设$\mu_{20}=\mu_{24}$进行检验.两样本的估计和假设检验问题中使用的理论和方法将在第十七章中讨论.

3.10 对照试验

对照试验(controlled experiment)一词在统计中有三个含义:(1)所有外部变量受到控制,(2)一个或多个**对照组**,或(3)由试验者控制的自变量水平.

外部变量是指所有**潜在**的可影响试验的自变量,但不是所研究的那个自变量.例如,3.9节中的鲑鱼生长试验可能会受到不同水平的食物、人类的触摸、生长空间、光照、水的化学成分等等的影响.当一个外部变量具有不同水平,这些水平对应于**真实自变量**(即试验者操纵的变量)的不同水平,则称它为**复合变量**.为了避免复合变量,我们可以尽量控制外部变量,使其在整个试验中尽可能保持在一个不变的水平.

即使是在最精密的正规试验室,对于外部变量的完全控制也几乎是不可能的,这只是一种理想情形.所以我们考虑通过以某种方式**随机选择**(见3.17节)**试验组**和**对照组**来抵消外部变量的作用.如果一个试验的不同组的研究主体是人或动物时,我们称之为**对象**;否则,称之为**试验单元**.

由于通常不可能得到整个总体,所以大多数试验都是在自然样本上进行的,在一个试验中,某个控制条件下得到的样本称为一个**试验组**.3.9节中的探索性试验有两个试验组,20℃的试验组和24℃的试验组.与此不同,对照组则提供**标准条件下的数据**,以便与一个或多个试验组(**处理组**)进行比较.例如,想确定某种新的感冒疫苗是否有效,我们从成人志愿者中随机抽取两组对象,其中一组注射带疫苗的生理盐水(试验组),另一组则注射不含疫苗的生理盐水(对照组),其他条件均相同.然后,在一段时间后通过比较两组对象中感冒者的人数可以确定疫苗的效果.探索性试验用于研究的假说形成阶段,而对照试验一般用于假设检验阶段(见3.8节).

对照试验的第三个含义是由试验者控制的自变量水平.试验者可以控制哪些对象或试验单元接受哪种处理,这正是对照试验和**观察式研究**(见3.11节)之间的主要区别.

3.11 观察式研究

如前面所讨论的,在试验中调查者通过改变自变量的值来研究其对因变量的影响.观察式研究则与之不同,在观察式研究中,调查者不能控制自变量,而只能通过观察存在的现象来考察自变量和因变量之间的关系.无论是在研究的探索性(假说形成)阶段还是假设检验阶段都可以进行观察式研究.

一个探索性观察研究实例是对统计学入门课程中成功学生的特征研究.如果成功用因变量——课程的总成绩来测量,则为了分离可能的影响因素,我们取的自变量有:每星期做家庭作业的时间,过去学过的数学课程数以及每门课的成绩,等等.分析变量间可能关系的最常用的统计方法称为**回归**和**相关**(见第十九章).

当出于伦理或实践的原因不能用其他的对照试验来检验一个现存的研究假说时,使用假设检验观察研究.为此,必须找到一个已经存在的**自然试验**(或**半试验**),在这个试验中,假设的自变量有不同的水平,因变量也有不同的水平.例如,要检验一种女性避孕药是否会导致血压的升高,我们可能研究两组人的血压(因变量),其中一组人使用这种避孕药(自变量),而另一组人不使用这种避孕药.

3.12 调查和普查

调查一般是指为了得到有关一组事物特征的信息所作的尝试,在统计中,调查特指通过提问从样本或总体中得到信息,例如市场调查,民意调查,电话调查等等.如果一个总体的所有元素都参与了调查,那么称这种调查为100%**调查**或**普查**.尽管调查和普查是观察式研究所使用的方法,但是它们也能用于试验中(例如调查观众在观看电影前后对电影的看法).

3.13　参数和非参数统计方法

适用于对给定数据集进行分析的推断方法是由一些复杂的因子所决定的. 前面已经讨论过这样一些因子,例如,试验中接受处理的组数及对照组的使用(见 3.9 节和 3.10 节);观察式研究中数据是探索性的还是假设检验的(见 3.11 节). 现在,我们考虑另外两个因子:测量总体的预期特征和数据的测量水平.

参数统计方法(简称为**参数统计**)基于对所研究总体和样本的特征的非常精确的和限制性的假设. 这些假设陈述了研究总体的预期特征(例如,总体参数的性质、总体分布的形状),并且指出抽取样本的类型,我们将这些假设称为**参数假设**. 如果这些假设能够满足,并且样本数据是间隔或比例水平的(见 2.6 节和 2.7 节),那么应该在推断分析中使用参数方法. 参数方法通常是首选的,这是因为:(1)对于统计假设,参数方法能提供最灵敏的检验;(2)参数方法能从数据中提取最多的信息;(3)参数方法能对最复杂的研究设计进行分析.

参数统计是最早提出的推断方法. 但是,人们很快意识到还需要一种等价的、使用更广泛的、假设条件更弱的推断方法,于是提出了**非参数统计方法**(简称为**非参数统计**). 我们称之为非参数是因为假设条件和统计假设都与参数无关. 一些非参数统计也被称为**分布自由统计**,这是因为它们不需要对总体的分布形状有任何特别的要求.

广义来说,非参数方法可以分为两类:用于名义水平数据的方法和用于次序水平数据的方法(见第二十章). 对于间隔水平和比例水平的数据,如果分析只限于数据的名义性质(=,≠)或者次序性质(<,>)(见 2.4 节和 2.5 节),那么也可以使用这些非参数方法. 对间隔水平和比例水平的数据总是推荐参数方法,所以只有参数假设严重背离时才选用非参数方法. 尽管那样,我们还是常常做数据变换使得可以使用参数分析.

图 3-3 给出了如何选择合适的分析方法的流程图. 对于间隔水平和比例水平的数据,如果参数假设条件能够满足,那么应使用参数方法进行分析;如果假设条件不能够满足,那么先尝试变换数据. 如果变换后假设条件能够满足,那么接着进行参数分析;否则,使用名义水平或次序水平的非参数方法对数据进行分析. 如果数据是名义水平或次序水平,那么采用合适的非参数方法. 在完成这些参数非参数决策之后,才开始考虑其他的决策因子(组数、对照组的使

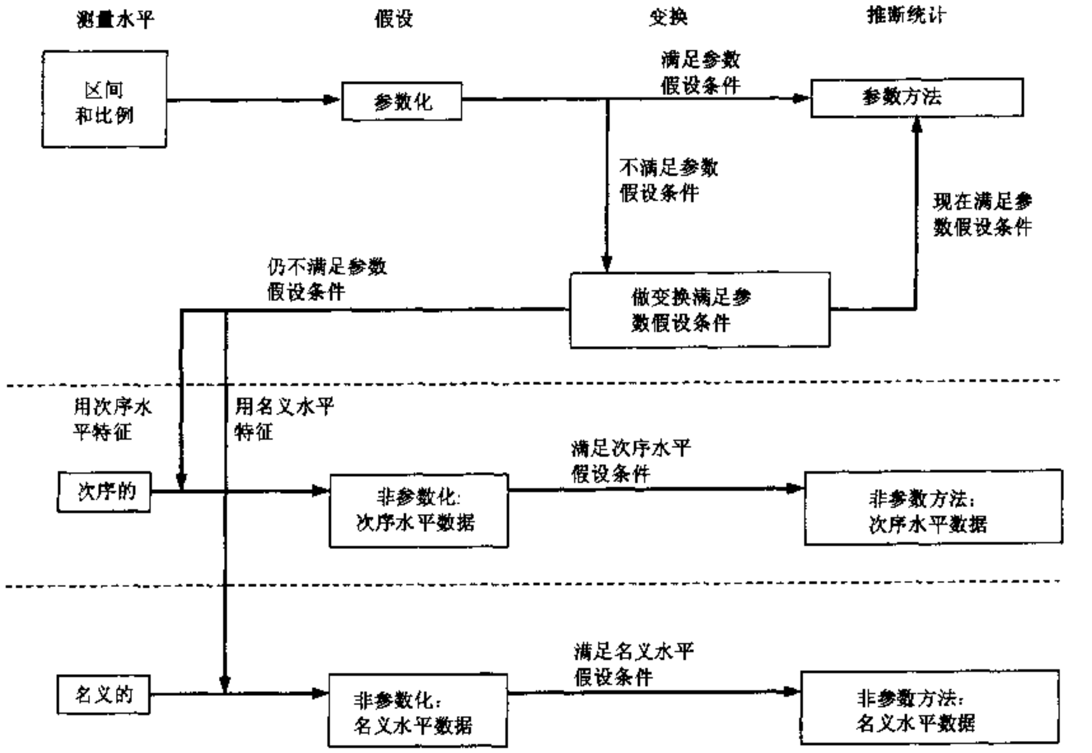


图 3-3

用、观察式研究是探索性的还是假设检验的).

3.14 数理统计学和一般统计学

数学的所有形式都源于一系列的假设(**公理**),基于这些假设经过演绎推理就得到**定理**和**完整的数学系统**.**数理统计学**(见1.1节)用到了多个数学领域(例如:概率论、微积分、高等代数)的基础,给出了完整的统计学体系.数理统计学是公理和定理紧密结合的一个完整的数学系统,而**一般统计学**(见1.1节)则给出了比较简单的水平.在一般统计学中,描述和推断方法来自数理统计学,但是几乎不解释这些方法是如何从数学体系中得到的.因为一般统计学不是从数学上讨论统计概念和方法,所以它是从**直觉水平**而非**数学水平**来表述的.本书是一本从直觉水平表述的有关一般统计学的书,理解这本书所需要的所有数学知识已在第一章中给出.

本书,以及其他有关一般统计学的书,具有不需要对各种方法进行数学推理的优点,但这也是缺点.要求学生接受的概念和方法都是直接给出的,并没有进行证明.因此,在整本书中或其他类似的直觉水平的书中,像"必须认为它是正确的"或"这可以从数学上证明"这样的句子很常见.只有在数理统计学的书中才会给出整个完整的系统.

3.15 抽样设计

到现在为止,我们已经介绍了推断统计的三种理论:概率论(见第八章到第十二章)、估计理论(见第十四章和第十五章)和假设检验理论(见第十六章).现在我们介绍推断统计的第四种理论:**统计抽样理论**(简称为**抽样理论**).这个理论处理测量总体和取自这个总体的测量样本之间的理论关系,将在第十三章中讨论.

抽样理论的一个方面是讨论推断统计的抽样过程——从总体中抽取合适样本的方法,使得由样本到总体的推断(见3.5节)是合理的.这一抽象的数学抽样过程称为**抽样设计**,即对一个存在的测量总体(称为**简单总体**),指定一种从中抽取测量样本(称为**简单样本**)的推断方法.图3-4给出了总体—样本—总体的理论次序.

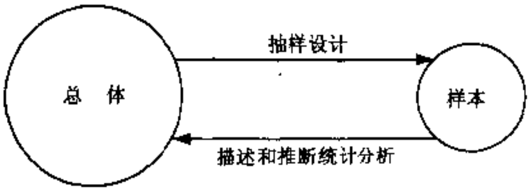


图3-4

抽样理论描述了一个纯粹的数学世界,然而实际抽样问题是在实际世界的复杂条件下解决的.为了解决这样一个问题,首先要定义一个感兴趣的自然总体;然后根据一个适当的理论抽样设计从中抽取一个自然样本,对自然样本进行测量产生一个测量样本;最后对数据进行描述和推断统计分析得到有关测量总体的推断.图3-1及3-2给出了这个过程.

下面我们对一些重要的理论抽样设计及它们的实际使用进行简要介绍.

3.16 抽样的概率:有放回和无放回

推断统计学的所有方面,包括抽样设计,都是基于概率理论的.有关概率论的知识将在第八章至第十二章中介绍,在这里,为了讨论抽样设计,我们对其中的一些概念作简要介绍.

一个事件的概率是指该事件将要发生的可能性,它是用(0)到(1)之间的一个数表示的.概率为(0)表示事件不可能发生,概率为(1)表示事件必定发生.例如,一个男子生小孩的概率为(0),明天太阳从东边升起的概率为(1).概率为(0.5)意味着事件将有一半的机会发生.

正如我们将在第八章中讨论的,解释和计算概率的方法有四种:**古典解释**、**相对频率解释**、**集合论解释**和**主观解释**.在这里,我们只考虑最古老、最常用的古典解释.

古典解释是在19世纪为研究赌博(掷骰子、抛硬币、抽纸牌等)中的机会问题而得到发展的.它可应用于任一符合下列条件的机会游戏:游戏可以在相同条件下一次又一次的重复进行,每次**试验**中,所有可能**结果**是已知的、互斥的,且发生的可能性相同.这些结果可以分成不同的类,称为**事件**.在一次试验中,事件A发生的概率用$P(A)$表示,可用如下公式计算

$$P(A)=\frac{m}{n} \tag{3.3}$$

其中m表示能使事件A发生的结果数,n表示所有可能的结果数.

例如,从一副已洗好的52张纸牌中抽取一张牌,我们要计算抽到A(爱司)的概率P(爱司).已知所有可能结果有52种——52张牌中的每一张牌都是一种可能结果,并且知道A(爱司)出现的结果有4种,应用式(3.3)得到:

$$P(\text{爱司})=\frac{4}{52}=\frac{1}{13}$$

如果在这次试验中抽到一张A(爱司),现在我们要计算再抽到一张A(爱司)的概率.进行第二次试验的方式有两种:(1)将这张A(爱司)放回牌中,重新洗牌,再进行抽取;(2)不将这张A(爱司)放回牌中,重新洗牌,再进行抽取.第一种方式称为**有放回抽样**,第二种方式称为**无放回抽样**.如果是有放回抽样,那么第二次抽到A(爱司)的概率是

$$P(\text{爱司})=\frac{1}{13}$$

但是,如果是无放回抽样,那么第二次抽到A(爱司)的概率是

$$P(\text{爱司})=\frac{3}{51}=\frac{1}{17}$$

如果被抽样的总体无限大($N=\infty$),那么不管抽样有无放回,我们都假定总体的组成在抽样中保持不变.在实际中,当样本容量(n)与总体容量(N)相比非常小时(通常n不超过N的5%),也可进行如上假定.对于有限总体,如一副扑克牌,N不是很大,我们都应考虑抽样时总体的变化.

3.17　随机抽样

大多数研究的目的是取得有关总体的一般真实情况.然而由于难以得到整个总体,所以需要从总体中抽取一个样本,然后运用统计方法,从样本信息推断关于总体某些特征的结论.为了使推断合理,样本必须在理论抽样设计(称为**随机抽样设计**,或**随机抽样**)的严格条件下抽取.

一个总体可以通过一次抽取一个或一组元素来进行抽样.抽取的基本单元称为**抽样单元**.在随机抽样中,总体中的每一个抽样单元都能以一已知的非零概率被抽到样本中去.由于被抽到的概率在抽样前是确定的,所以随机抽样也称为**概率抽样**.为了保证这个指定的概率,在选取过程中需要引入**随机性**(或机会),这是通过**随机数表**(见3.23节)或在计算机及许多计算器上用**随机数生成函数**产生的数列随机确定的.

在理想的理论情形中,所研究总体(称为**目标总体**)的所有抽样单元是已知的,可以用一张表来记录,这张表称为**抽样框架**.然后选择一个随机抽样设计以指定抽样框架中所有抽样单元被抽中的概率,最后按照指定概率,使用某种随机选择方法抽取样本.

下面介绍四种重要的随机抽样设计:简单随机抽样、分层随机抽样、系统随机抽样和整群随机抽样.

3.18　简单随机抽样

简单随机抽样是一个数学概念,将在讨论概率论后的第十三章详细讨论.这里,只给出它的非数学的、直观的(见3.14节)定义:

简单随机抽样是使总体中所有抽样单元都有相等的概率被抽取到样本中去的一种抽样方

法. 用这种方法得到的样本称为**简单随机样本**.

考虑一个 35 个人构成的目标总体,从中抽取一个 3 人的简单随机样本. 规定抽样单元是单个人,我们将这 35 个人的名字分别写在完全相同的纸上,并将这些纸放在一个碗中,闭上眼睛,无放回地一次抽取一张纸,总共抽取三张纸. 这样就得到一个简单随机样本,因为所有的名字都有相等的被抽取的概率,第 1 次是 1/35,第 2 次是 1/34,最后一次是 1/33.

初等推断统计的大部分方法基于如下假设:样本是通过简单随机抽样得到的,因此是简单随机样本. 因为这是个非常基本、常用的假定,所以**本书的剩余部分,如未进行特别说明,样本或随机样本都是指简单随机样本**.

3.19 分层随机抽样

为了由样本到总体的推断有效,样本必须是总体的代表. 简单随机抽样通常能提供具有代表性的样本,但是有时**分层随机抽样**能提供更具代表性的样本,如果总体包含一些不重叠的互斥的部分(称为**层**),这是由年龄、性别、种族及地理位置等因子引起的. 为使用分层随机抽样,要求层有相对同质性,即各层间的差异比层内间的差异大的多. 如果这样的层已经存在,那么分层随机抽样设计从每一层抽取一个随机样本[简单抽样,系统抽样(见 3.20 节),整群抽样(见 3.21 节)或其他]. 若每一层的随机样本容量在样本中所占的比例与该层元素在总体中所占的比例(层的大小/总体大小)相等,则称为**比例分层随机抽样**. 否则,称为**非比例分层随机抽样**.

考虑分层随机抽样问题:一群男性中,5%为高收入者、65%为中等收入者、30%为低收入者,试确定他们的平均收入. 将总体按照收入水平分为三层,若进行比例分层随机抽样,则在样本中有 5%的高收入者,65%的中等收入者,30%的低收入者;若进行非比例分层随机抽样,那么每一层的百分比可以任意确定,例如,用相等的百分比 $33\frac{1}{3}\%$.

3.20 系统随机抽样

如果抽样框架非常大,通常采用**系统随机抽样**. 在系统随机抽样中,抽样框架中每相隔 k 的元素都被取为样本元素,而第一个被抽取元素(称为**初始元素**)是从前 k 个元素中随机选择的. 例如,如果每隔 50 抽取一个元素,那么为了确定初始元素,从数 1—50 中随机抽取一个数的简单随机样本,假设抽取得到数 20,那么框架中第 20,第 70,第 120 等等均包含在样本中. 这个样本是随机样本,因为总体中的每个单位被抽到的概率都是事先知道的$\left(\frac{1}{k}=\frac{1}{50}\right)$.

如果在抽样框架中存在明显的周期性或循环,应避免使用系统随机抽样. 例如,在调查住宅的过程中,抽样间隔不要取为 25,因为一个街区的第 25 栋住宅一般处在拐角处,是该街区中最贵的住宅.

3.21 整群随机抽样

如果一个总体很大且非常分散,那么**整群随机抽样**能使抽样成本相对较低. 将总体元素划分成若干互斥部分(称为**群**),每部分尽可能是异质的(与分层随机抽样的同质层不一样). 在**单阶段整群抽样**中,随机抽取 n 个群,被抽到的群中元素均包括在样本中. 在**两阶段整群抽样**中,首先随机抽取 n 个群,然后在抽到的每一个群中再进行第二次随机抽样. 如果抽样过程包括两个以上抽样阶段,则称之为**多阶段整群抽样**.

例如,对纽约这样一个大州的选举进行民意调查,纽约州的每一个郡,由于其相对异质性,可以看作抽样中的一个群. 若进行单阶段随机抽样,则随机抽取若干个郡,郡中的每一个合法选举人都作为样本. 然而更常用的为多阶段随机抽样,即先随机抽取若干个郡,然后在被选中的郡中抽取若干个区,以此类推,一直抽到最后一个阶段为止.

3.22 非随机抽样

我们在 3.17 节中提到：在随机抽样中，总体中的每一个抽样单元都能以一已知的非零概率被抽到样本中去. 不满足这一抽样原则的抽样设计称为**非随机抽样设计**（或**非概率抽样**）. 因为它们与调查者关于哪些抽样单位应包括在样本中的个人的、非随机的判断有关，所以也称为**判断抽样设计**. 这种抽样一般会导致某种**抽样偏差**（系统抽样误差），可能是**选择偏差**，也可能是**响应偏差**（测量的系统误差也称为偏差，见 2.13 节），所以也称为**有偏抽样设计**. 如果抽样过程倾向于从总体中选择确定的元素，而排斥另一些元素，就会导致选择偏差. 如果抽样设计导致大量数据缺失，就产生了响应偏差. 例如，邮寄的问卷调查表的返回率很低，就会产生响应偏差.

例 3.4 下例是一个既有选择偏差也有响应偏差的非随机抽样，请说明原因.

美国 1936 年总统选举的两个候选人是共和党人阿尔佛雷德 · 兰德勒（Alfred M. Landon）和在任总统——民主党人富林克林 · 罗斯福（Franklin D. Roosevelt）. 选举前几个星期，Literary Digest 杂志想预测选举结果. 该杂志根据杂志订阅名单、电话姓名地址目录以及汽车登记记录，给 1000 万个选举人邮寄了问卷调查表. 返回的问卷调查表约有 230 万份，其中有 57% 的选举人支持兰德勒. 于是根据这些调查结果，该杂志预测兰德勒将获胜. 然而，几个星期后，在选举中获胜的却是罗斯福，他的得票率是 62%.

解 ☞ 因为在上述抽样中，抽样局限于杂志订阅者以及电话、汽车的拥有者，使得大多数选举人被抽到样本中的概率为 0，所以这是一个非随机抽样. 1936 年的美国正处在经济萧条时期，这一判断选择将样本局限于一个相对富有的阶层. 目标总体和抽样框架之间的差异导致了抽样的选择偏差. 此外，由于只有 25% 的被调查人返回了问卷调查表，因此存在响应偏差，也称为**自选择偏差**. 即使对这些被抽取到的阶层，在抽样前它们被抽到样本中的概率也是未知的.

3.23 随机数表

将机会引入抽样过程的最常用工具是**随机数表**（或**随机数字表**）. 随机数表一般是通过计算机的随机数生成函数产生的，它由数千个数字组成，每一个数字都是 0—9 之间的任一个数. 事实上，表中的每一个数字都是从 0 到 9 中选出的一个简单随机样本. 因此，0—9 之间的每一个数都有相等的可能性出现在表中的任一个数字位置上，并且表中数字之间没有系统的联系. 附录中的表 A.1（随机数）就是一个随机数表，它是由 6000 个数字以两位数字为一组排列而成，总共有 40 行 75 列.

例 3.5 图 3-5 是统计学班级 64 个学生的座位表. 表中的每个中括号（[ ]）表示一个学生的座位，括号中给出了学生的名字首字母和性别（f 表示女性，m 表示男性）. 班级中有 16 个女生（25%）和 48 个男生（75%）. 请用附录中的表 A.1 从该班选择一个容量为 16 的简单随机样本.

	1	2	3	4	5	6	7	8
1	[CA-m]	[FE-m]	[LB-f]	[HE-m]	[LW-m]	[OA-m]	[PS-m]	[OF-m]
2	[AA-f]	[HC-m]	[EB-m]	[MA-m]	[ME-m]	[HK-m]	[AD-m]	[RE-m]
3	[AE-f]	[FA-m]	[CE-m]	[BP-m]	[EO-m]	[RA-m]	[DA-m]	[GK-m]
4	[JA-m]	[GB-m]	[MJ-f]	[NO-f]	[JW-m]	[LT-f]	[HO-m]	[WA-m]
5	[JD-f]	[DD-f]	[NA-m]	[SM-m]	[MQ-m]	[JT-m]	[TS-m]	[MU-m]
6	[GM-m]	[BC-m]	[CI-m]	[EF-f]	[JL-m]	[JQ-m]	[FV-m]	[DW-m]
7	[AC-f]	[DM-m]	[JH-f]	[BF-m]	[AH-f]	[NP-m]	[GT-m]	[GY-f]
8	[MZ-f]	[BJ-f]	[LR-m]	[CR-m]	[PB-m]	[EJ-m]	[AT-m]	[TM-f]

图 3-5

解 ☞　为了得到这样的样本,必须满足随机抽样(见 3.17 节)和简单随机抽样(见 3.18 节)的条件.将班级看作是一个总体,并用 01—64 对这些学生依次编号,从而产生一个抽样框架(见图 3-6).然后,我们根据随机数表 A.1 无放回地抽取学生号码,第一次抽取时所有学生都有$\frac{1}{64}$的概率被抽中;第二次抽取时所有剩余学生都有$\frac{1}{63}$的概率被抽中;如此继续下去,第十六次抽取时所有剩余学生被抽中的概率是$\frac{1}{49}$.我们只需从表 A.1 选择 16 个 01—64 之间的无重复随机数,就完成了这个抽样.

	1	2	3	4	5	6	7	8
1	[CA-m] 01	[FE-m] 02	[LB-f] 03	[HE-m] 04	[LW-m] 05	[OA-m] 06	[PS-m] 07	[OF-m] 08
2	[AA-f] 09	[HC-m] 10	[EB-m] 11	[MA-m] 12	[ME-m] 13	[HK-m] 14	[AD-m] 15	[RE-m] 16
3	[AE-f] 17	[FA-m] 18	[CE-m] 19	[BP-m] 20	[EO-m] 21	[RA-m] 22	[DA-m] 23	[GK-m] 24
4	[JA-m] 25	[GB-m] 26	[MJ-f] 27	[NO-f] 28	[JW-m] 29	[LT-f] 30	[HO-m] 31	[WA-m] 32
5	[JD-f] 33	[DD-f] 34	[NA-m] 35	[SM-m] 36	[MQ-m] 37	[JT-m] 38	[TS-m] 39	[MU-m] 40
6	[GM-m] 41	[BC-m] 42	[CI-m] 43	[EF-f] 44	[JL-m] 45	[JQ-m] 46	[FV-m] 47	[DW-m] 48
7	[AC-f] 49	[DM-m] 50	[JH-f] 51	[BF-m] 52	[AH-f] 53	[NP-m] 54	[GT-m] 55	[GY-f] 56
8	[MZ-f] 57	[BJ-f] 58	[LR-m] 59	[CR-m] 60	[PB-m] 61	[EJ-m] 62	[AT-m] 63	[TM-f] 64

图 3-6

使用随机数表 A.1 的第一步是在随机表中确定一个两位数字作为**初始位置**,选择初始位置的方法有很多,例如:闭上眼睛,在表中任指一点,取最近的两位数作为初始位置的列数;用同样方法得到初始位置的行数.按照这种方法,我们得到初始位置的列数是 59,行数是 24.处在此位置的数是 68.

我们可以从初始位置出发向任一方向(向上、向下、对角方向、向左或向右)选取两位的随机数.这里我们是从初始位置向下选取 01—64 之间的无重复随机数.一直到第 59 列的底端,选到的随机数为:59,64,17,22,07,39,44,32,26,53,45,38 和 13.这里只有 13 个数.为了得到剩余的 3 个数,接着从第 60 列的顶端开始向下选择,得到 63,27 和 37.根据这些随机数得到容量为 16 的简单随机样本为:

[LR-m],[TM-f],[AE-f],[RA-m],[PS-m],[TS-m],[EF-f],[WA-m],
[GB-m],[AH-f],[JL-m],[JT-m],[ME-m],[AT-m],[MJ-f],[MQ-m].

习题解答

总体

3.1 指出下列总体是有限总体还是无限总体,是自然总体还是测量总体:(a) 游过英吉利海峡的所有还活着的女子的当前年龄,(b) 为了让一只挪威鼠在特定条件下学会走某个迷宫所进行的试验次数,(c) 一个 20 岁男子的所有的体细胞,(d) 按照某个特定配方烹制的所有巧克力蛋糕,(e) 所有现在和过去的美国公民的体重.

解 ☞　(a) 有限测量总体

(b) 无限测量总体

(c) 有限自然总体

(d) 无限自然总体
(e) 有限测量总体

样本

3.2 指出下列样本是自然样本还是测量样本:(a) 伦勃朗的四幅油画的表面积,(b) 四分钟内跑完了一英里路程的六个还活着的男子,(c) 一个湖的日水温(℃)的 20 天测量值.

解 ☞ (a) 测量样本
(b) 自然样本
(c) 测量样本

3.3 对下列各项,首先指出它们是总体还是样本或者两者都是,然后指出是测量的还是自然的:(a) 所有健在的美国前总统现在的体重,(b) 50 个美国人对移民的态度,用数字 1 到 5 度量,其中 1=不欢迎,…,5=非常欢迎,(c) 一个连锁商店的每个分店在一个月内售出的新电视机的台数,(d) 濒临灭绝的某个蝴蝶种类中仅存的 132 只蝴蝶.

解 ☞ (a) 总体,测量
(b) 样本,测量
(c) 两者都是,测量
(d) 总体,自然

3.4 什么时候需要用自然样本代替自然总体?

解 ☞ 尽管对任何一个研究领域在进行数据收集和分析时总是希望得到整个自然总体的特征,但是要测量整个总体是不大实际或不大可能的.例如,要全部测量一个无限自然总体是不可能的(见 3.2 节).即使对一个有限总体,如果该总体非常大,或者在时间或空间上的跨度很大,那么要进行全部测量也是不可能的.最后,即使一个自然总体是有限的并且是可测量的,但是如果测量过程是具有破坏性的,那么也不能对整个总体进行测量.例如,若一个苹果种植者测量他的所有苹果的含糖量,那么他将没有苹果可供销售.所以在大多数情形下,总是从一个总体中抽取自然样本进行测量,然后运用推断统计学的方法将样本推广到得不到的总体.

参数与统计量

3.5 计算算术平均数:(a) 测量样本 $x_1=4$, $x_2=2$, $x_3=3$,(b) 测量总体 $x_1=5$, $x_2=5$, $x_3=5$.

解 ☞ (a) $\bar{x}=\dfrac{\sum_{i=1}^{n}x_i}{n}=\dfrac{\sum_{i=1}^{3}x_i}{3}=\dfrac{4+2+3}{3}=3$

(b) $\mu=\dfrac{\sum_{i=1}^{N}x_i}{N}=\dfrac{\sum_{i=1}^{3}x_i}{3}=\dfrac{5+5+5}{3}=5$

估计问题和假设检验问题

3.6 为什么推断统计学也称为**归纳统计学**?

解 ☞ 逻辑学领域将人类的推理形式分成两大类:**归纳推理**(也称为**归纳逻辑**)和**演绎推理**(也称为**演绎逻辑**).归纳推理就是一般化的过程.下面是一个归纳推理的实例:

所有母牛都是哺乳动物并且有大脑
所有人类都是哺乳动物并且有大脑
所有狗都是哺乳动物并且有大脑

所以,**可能**所有哺乳动物都有大脑

这个论证是由具体例子来得到一般(或普遍的)结论,这是归纳推理的特征.在结论中,"可能"一词表明这是个不确定的结论,仅仅是一个猜测.结论的本质是:由我们所知的,"所有哺乳动物都有大脑"可能是对的.

演绎推理与归纳推理正好相反.它是从一般结论到具体例子.下面是一个演绎推理的实例:

所有哺乳动物都有大脑

人类是哺乳动物

所以,人类有大脑

这个演绎论证是从一般陈述(称为**前提**)推到一个具体的结论.如果这个前提是真实的,那么得到的具体结论就一定是真实的.由演绎推理得到的结论不是一个不确定的可能性的结论,而是一个完全确定的结论.演绎推理研究前提意味着如果前提为真,那么结论也为真.

估计问题和假设检验问题中的推断统计使用的都是归纳推理.它根据具体的有限的样本信息推出有关整个测量总体的结论(统计推断),并且这些结论是不确定的概率陈述.然而,与非统计学归纳结论不同的是,统计学归纳还给出了有关该结论真实程度的定量估计.

试验和观察式研究

3.7 指出下列哪些是探索性试验,哪些是对照试验,哪些是观察式研究:(a) 调查美国1500名合法选举人,在下届总统选举中他们支持哪位候选人,(b) 通过对两个医院的外科医生的手术计时,确定每个医院作阑尾切除手术的平均时间,(c) 随机选取两组女性,其中一组服用维生素C,另一组则不服用维生素C,其他条件均相同.然后,通过比较两组对象中在一段时间内感冒者的人数来确定维生素C是否能预防感冒.

解 ☞ (a) 观察式研究

(b) 观察式研究

(c) 对照试验

3.8 指出下列哪些是探索性试验,哪些是对照试验,哪些是观察式研究:(a) 通过对两个国家的新生婴儿测量体重确定两国家新生儿的平均体重,(b) 将一种新的杂交花种植在四种氮含量不同的土壤中确定适合生长的氮的最优水平,(c) 为了检验一种新型的汽油添加剂是否真的能提高每加仑的英里数,随机选取两组汽车,其中一组使用这种汽油添加剂,另一组则不使用,其他条件均相同,比较两组汽车的每加仑的英里数的平均数.

解 ☞ (a) 观察式研究

(b) 探索性试验

(c) 对照试验

抽样的概率:有放回和无放回

3.9 请根据公式(3.3),计算(a) 第一次抛硬币正面向上的概率,(b) 第二次抛硬币正面向上的概率?

解 ☞ (a) $\frac{1}{2}$, (b) $\frac{1}{2}$

3.10 一副标准的扑克牌(52张)中有13张红桃牌,从中抽一张牌,请根据公式(3.3),计算(a) 抽到红桃10的概率,(b) 抽到任一张红桃牌的概率

解 ☞ (a) $\frac{1}{52}$

(b) $\frac{13}{52}=\frac{1}{4}$

3.11 在习题3.10中,如果第一次抽到一张红桃,那么第二次又抽到一张红桃的概率是多少:(a) 采取有放回抽样,(b) 采取无放回抽样?

解 ☞ (a) $\frac{13}{52}=\frac{1}{4}$

(b) $\frac{12}{51}=\frac{4}{17}$

随机抽样与非随机抽样

3.12 某班级按照性别分成了两个层,其中女生占25%,男生占75%,见图3-6.对这些层使

用随机数表 A.1 和简单随机抽样方法，从该班抽取一个容量为 16 的比例分层随机样本(见 3.19 节).

解 ☞　若进行按性别的比例分层随机抽样，则在女生中抽取容量为 16 的样本的 25%，在男生层中抽取样本的 75%，即从女生中抽取 4 个，从男生中抽取 12 个. 对每一层分别进行简单随机抽样来得到这些样本. 对女生，先使用例 3.5 中的方法确定一个初始位置(37 列和 40 行的交点处:42)，然后从初始位置开始在 37 列向上选取得到 4 个不重复的女生号码:49,17,27,51. 对 12 个男生，得到一个初始位置(30 列和 29 行的交点处:34)，然后从这开始在 30 列向下，如果必要，由 31 列底端向上选取。这样得到下列不重复的男生号码:06,48,42,29,41,55,46,52,43,07,47 和 01. 因此我们的容量为 16 的比例分层随机样本为:

[AC-f],[AE-f],[MJ-f],[JH-f],[OA-m],[DW-m],[BC-m],[JW-m],
[GM-m],[GT-m],[JQ-m],[BF-m],[CI-m],[PS-m],[FV-m],[CA-m].

3.13　将图 3-5 中的学生随意的划分成每四个学生一组的 16 个群，见图 3-7. 对这 16 个群使用随机数表 A.1 和简单随机抽样方法，从该班级抽取一个容量为 16 的单阶段整群随机样本(见 3.21 节).

	1	2	3	4	5	6	7	8
1	[CA-m]	[FE-m]	[LB-f]	[HE-m]	[LW-m]	[OA-m]	[PS-m]	[OF-m]
	01		02		03		04	
2	[AA-f]	[HC-m]	[EB-m]	[MA-m]	[ME-m]	[HK-m]	[AD-m]	[RE-m]
3	[AE-f]	[FA-m]	[CE-m]	[BP-m]	[EO-m]	[RA-m]	[DA-m]	[GK-m]
	05		06		07		08	
4	[JA-m]	[GB-m]	[MJ-f]	[NO-f]	[JW-m]	[LT-f]	[HO-m]	[WA-m]
5	[JD-f]	[DD-f]	[NA-m]	[SM-m]	[MQ-m]	[JT-m]	[TS-m]	[MU-m]
	09		10		11		12	
6	[GM-m]	[BC-m]	[CI-m]	[EF-f]	[JL-m]	[JQ-m]	[FV-m]	[DW-m]
7	[AC-f]	[DM-m]	[JH-f]	[BF-m]	[AH-f]	[NP-m]	[GT-m]	[GY-f]
	13		14		15		16	
8	[MZ-f]	[BJ-f]	[LR-m]	[CR-m]	[PB-m]	[EJ-m]	[AT-m]	[TM-f]

图 3-7

解 ☞　为了得到单阶段整群随机样本，首先用简单随机抽样方法从这 16 个群中抽取 4 个群，然后将被抽中的群中的所有学生都作为样本. 采用例 3.5 中的方法，首先确定一个初始位置(3 列和 6 行的交点处:57)，然后从初始位置开始在 6 行向右选取 4 个 01 到 16 之间的不重复的数. 一直到第 6 行的最右端，得到两个随机数:07 和 11. 为了选取另外两个数，从第 7 行的最左端(即第 1 列)开始继续向右，选取得到 06 和 01. 选取第 1 个、第 6 个、第 7 个和第 11 个群中的所有学生作为样本，得到单阶段整群随机样本为:

[CA-m],[FE-m],[AA-f],[HC-m],[CE-m],[BP-m],[MJ-f],[NO-f],
[EO-m],[RA-m],[JW-m],[LT-f],[MQ-m],[JT-m],[JL-m],[JQ-m].

3.14　一个心理学家想做一个老鼠学走迷宫的试验，他从一个动物销售处购买了同一种类的 20 只雄性老鼠，请问这些老鼠是否是简单随机样本?

解 ☞　从纯粹的数理统计学角度来说，这不是一个简单随机样本. 因为如 3.17 节及 3.18 节指出，简单随机样本需要满足以下两个条件:(1)总体中的每一个抽样单元(这里是单个雄性老鼠)都能以一已知非零概率被抽到样本中去;(2)每次抽样中，所有剩下抽样单元被抽到样本中的概率相等. 而上述老鼠总体及其他容量很大、很分散的总体明显不满足这些条件. 然而，在应用统计学中，我们常对这样的总体如此抽样，并认为这样的样本满足推断统计学的简单随机抽样的条件. 为什么?

数学模型的严格的抽象性质与对实际复杂数据做统计分析的实践需要之间存在着矛盾，而这个

例子就是我们将在此书中提到的众多例子中的一个.在这里,除非样本明显不具代表性[例如:例 3.4 中的目标总体与抽样框架之间的巨大差别;某种系统偏差的出现(如商人只提供晚年的老鼠);等等],否则样本被认为是简单随机样本.幸运的是,大多数统计模型是"稳健的",也就是说,即使违背了假设条件,只要它在一定的范围内,那么推断方法仍然能给出有效的结果.

补充习题

总体

3.15 指出下列总体是有限总体还是无限总体,是自然总体还是测量总体:(a) 登上过月球的所有男子,(b) 所有患有糖尿病的女子的血糖含量.

答案 (a) 有限自然总体,(b) 有限测量总体

3.16 指出下列总体是有限总体还是无限总体,是自然总体还是测量总体:(a) 所有圆的直径(以英寸为单位),(b) 抛硬币的结果(正面向上或反面向上).

答案 (a) 无限测量总体,(b) 无限测量总体

样本

3.17 对下列各项,首先指出是总体还是样本或者两者都是,然后指出是测量的还是自然的:(a) 某核工厂自建造以来,对方圆 50 里之内出现的所有新的癌症病例进行分类,(b) 某出租车公司的所有出租车一年内行走的英里数/加仑/月.

答案 (a) 两者都是,测量,(b) 两者都是,测量

3.18 对下列各项,首先指出是总体还是样本或者两者都是,然后指出是测量的还是自然的:(a) 某果园中的所有苹果,(b) 某个心理学试验中的 10 只白鼠.

答案 (a) 两者都是,自然,(b) 样本,自然

参数和统计量

3.19 计算下列样本或总体的算术平均值:(a) 测量样本 $x_1=0$、$x_2=0$,$x_3=3$,$x_4=5$;(b) 测量总体 $x_1=10{,}000$,$x_2=5{,}000$.

答案 (a) 2,(b) 7,500

试验和观察式研究

3.20 指出下列哪些是探索性试验,哪些是对照试验,哪些是观察式研究:(a) 确定非节假日上午 8 点至 10 点,经过某超市的可能选址的平均车辆数,(b) 为了确定某种新型"增强记忆的药物"的药效,比较两组老鼠学走迷宫的表现,其中一组在学走迷宫前注射含有此药物的生理盐水,另一组在学走迷宫前注射不含此药物的生理盐水,(c) 确定每个美国家庭的成人数.

答案 (a) 观察式研究,(b) 对照试验,(c) 观察式研究

3.21 指出下列哪些是探索性试验,哪些是对照试验,哪些是观察式研究:(a) 比较 20 岁的男子和女子平均肺活量的大小,(b) 对用于不同班级的三种新阅读教学方法的有效性进行检验,(c) 确定某个公司的股东们对于合并提议的态度.

答案 (a) 观察式研究,(b) 探索性试验,(c) 观察式研究

抽样的概率:有放回和无放回

3.22 在标准 52 张的一副纸牌中有一半是红色的牌,运用公式(3.3),确定从这副纸牌中抽到如下纸牌的概率:(a) 红色的牌,(b) 黑色的牌.

答案 (a) $\frac{26}{52}=\frac{1}{2}$,(b) $\frac{26}{52}=\frac{1}{2}$

3.23 在上题中,若抽到的是红色的牌,则运用公式(3.3),确定用如下抽样方法再抽取时抽到黑色牌的概率:(a) 有放回抽样,(b) 无放回抽样

答案 (a) $\frac{26}{52}=\frac{1}{2}$,(b) $\frac{26}{51}$

随机抽样与非随机抽样

3.24 使用随机数表 A.1 和简单随机抽样确定初始元素，从该元素开始，以 4 为间隔，从图 3-6 中抽取容量为 16 的系统随机样本.

答案 运用例 3.5 中的方法，首先确定一个初始位置(列数是 16，行数是 26，处在此位置的数是 65)，然后从初始位置开始向下选取得到 01 到 04 之间的一个数是 03，即初始元素的编号. 这表示在图 3-6 的抽样框架中，我们要抽取的样本的编号为：03，07，11，15，19，23，27，31，35，39，43，47，51，55，59，63. 所以得到系统随机样本为：

[LB-f]，[PS-m]，[EB-m]，[AD-m]，[CE-m]，[DA-m]，[MJ-f]，[HO-m]，
[NA-m]，[TS-m]，[CI-m]，[FV-m]，[JH-f]，[GT-m]，[LR-m]，[AT-m].

3.25 某大城市的一个电台脱口秀主持人在选举前通过热线电话调查选举人对市长的两个主要候选人的支持程度. 在打进热线电话的 800 个听众当中，有 500 人支持 A，250 人支持 B，50 人无所谓. 请问这是个随机样本吗?

答案 这样的电话调查不能认为是随机抽样，因为这样的样本不具有代表性. 脱口秀主持人只邀请了他的听众参加调查，这是一种判断选择，由于他把大多数潜在的选举人排除在外，所以产生了选择偏差. 此外，该抽样还存在响应偏差，这是因为只有感兴趣的听众才打进热线电话. 有的电话调查也能产生随机样本，例如：随机数拨号调查. 这种调查是从电话号码本中选择随机样本，并对选择的号码不断拨打直到有人接听为止.

第四章　描述性统计：将统计资料整理成表格形式

4.1 阵列和极差

将一组测量值按升序或降序排列的结果就称为一个**阵列**.在**升序阵列**中，测量值从小到大排列；在**降序阵列**中，测量值则从大到小排列.

一组测量值的**极差**是指最大测量值与最小测量值之差.若有变量 X 的一些测量值，那么

$$R = x_l - x_s \tag{4.1}$$

其中 R 表示极差，x_l 表示最大值，x_s 表示最小值.如果测量值已经表示成一个升序阵列，那么 x_s 和 x_l 就分别是阵列中的第一个和最后一个测量值.如果测量值已经表示成一个降序阵列，那么 x_s 和 x_l 的次序正好颠倒过来.

例 4.1　下面是一个长度测量值样本：5.1mm，2.9mm，6.4mm，9.2mm，7.7mm.(a) 将该样本表示成一个升序阵列，(b) 计算极差.

解 ☞　(a) 2.9mm，5.1mm，6.4mm，7.7mm，9.2mm

(b) 极差＝9.2mm－2.9mm＝6.3mm

例 4.2　附录中的表 A.2 是例 3.5 介绍的某统计学专业班级的多种测量的结果.请将第 4 列中 16 个女生的身高测量值排列成降序阵列，并计算极差.

解 ☞　16 个女生的身高(精确到 1/4 英寸)分别为：67.75，60.25，63.75，65.25，62.00，63.50，65.25，65.50，65.25，64.75，67.00，64.25，69.25，66.25，63.00，64.75.这些测量值的降序阵列为：69.25，67.75，67.00，66.25，65.50，65.25，65.25，65.25，64.75，64.75，64.25，63.75，63.50，63.00，62.00，60.25.极差为：R＝69.25 in－60.25 in＝9.00 in.

4.2 频数分布

频数分布是指测量尺度中的类在某测量集合中出现次数的汇总，它说明测量值在测量尺度上的**分布情况**.频数分布可以用表格表示，称为**频数表**，也可以作成各种图形形式(见第五章).

例 4.3　将下列 20 个重量测量值(以千克为单位)样本排列成升序阵列，并作出频数分布：1.0，1.5，1.3，1.3，1.3，1.4，1.4，1.0，1.2，1.2，1.3，1.2，1.3，1.3，1.4，1.3，1.5，1.2，1.4，1.3.

解 ☞　升序阵列为：1.0，1.0，1.2，1.2，1.2，1.2，1.3，1.3，1.3，1.3，1.3，1.3，1.3，1.3，1.4，1.4，1.4，1.4，1.5，1.5.将其频数分布表示成频数表，见表 4.1.

表 4.1

重量(kg) x_i	计数	f_i
1.0	||	2
1.1		0
1.2	||||	4
1.3	卌|||	8
1.4	||||	4
1.5	||	2
$\sum$		20

频数表有三列：**重量**、**计数**和**频数**.重量这一列给出了测量尺度中用到的所有的类，其中 x_i 表示变量 X(重量)的第 i 个类.这里总共有 6 个类，从 $x_i = x_1 = 1.0$ 到 $x_k = x_6 = 1.5$.

计数列给出了每一个类中测量值个数的**计数**，这是由**计数标志**完成的.计数标志有多种，但最常用的是用简单的竖线表示单个观测值(见表 4.1)，通常用 4 个竖线加 1 个对角线表示 5 个测量值.因此，列中的每一个竖线表示测量值中对应的类出现一次.

频数列是计数列的数值表示.记号 f_i 表示每一个类出现的频数.

如果测量集合是一个样本，这时频数的合计等于样本容量 n，即

$$\sum_{i=1}^{k} f_i = n$$

其中 k 表示类的个数. 如果测量集合表示一个整个总体，那么频数的合计等于总体容量 N，即

$$\sum_{i=1}^{k} f_i = N$$

4.3 相对频数分布和百分数分布

将每一个类的频数除以样本容量就得到该类在样本中的**相对频数**(**或比例**)，记为 f_i/n. 对总体而言，相对频数则是每一个类的频数除以总体容量，记为 f_i/N. 将相对频数表示成百分数形式就得到类的百分数. 对于样本，记为 $[(f_i/n)\times 100]\%$，对于总体，记为 $[(f_i/N)\times 100]\%$.

相对频数分布和**百分数分布**用来说明相对频数和百分数在测量尺度上的分布情况. 与频数分布一样，它们也可用表格或各种图形(见第五章)来表示.

例 4.4 在表 4.1 中增加表示相对频数和百分数的列.

解 ☞ 表 4.1 中增加表示相对频数和百分数的列之后的结果见表 4.2.

注：表 4.2 中并未出现表 4.1 中的计数列，这是因为该列仅仅在频数分布的生成过程中有用，在统计信息汇总中是很少出现的.

表 4.2

重量(kg) x_i	频数 f_i	相对频数 f_i/n	百分数 $[(f_i/n)\times 100]\%$
1.0	2	2/20=0.1	10
1.1	0	0/20=0.0	0
1.2	4	4/20=0.2	20
1.3	8	8/20=0.4	40
1.4	4	4/20=0.2	20
1.5	2	2/20=0.1	10
$\sum$	20		

例 4.5 在波士顿马拉松比赛中，前 30 名男运动员的成绩(单位：分；连续比例测量)的升序阵列如下：129，130，130，133，134，135，136，136，138，138，138，141，141，141，142，142，142，142，143，143，143，143，143，144，144，145，145，145，145，145. 将此阵列汇成一个表格，各列为：时间、频数、相对频数和百分数.

解 ☞ 汇总表见表 4.3.

表 4.3

时间(min) x_i	频数 f_i	相对频数 f_i/n	百分数 $[(f_i/n)\times 100]\%$	时间(min) x_i	频数 f_i	相对频数 f_i/n	百分数 $[(f_i/n)\times 100]\%$
129	1	0.033	3.3	138	3	0.100	10.0
130	2	0.067	6.7	139	0	0.000	0.0
131	0	0.000	0.0	140	0	0.000	0.0
132	0	0.000	0.0	141	3	0.100	10.0
133	1	0.033	3.3	142	4	0.133	13.3
134	1	0.033	3.3	143	5	0.167	16.7
135	1	0.033	3.3	144	2	0.067	6.7
136	2	0.067	6.7	145	5	0.167	16.7
137	0	0.000	0.0	$\sum$	30		

4.4 分组频数分布

将表 4.3 中的时间按每 3 分钟一组压缩成六个组，第一组 128—130，从最快成绩的前一分钟开始，再往后两分钟；下一组从 131 到 133 分钟；直到最后一个，从 143 到前 30 名中的最慢成绩 145 分钟，就得到一个**分组频数分布**，见表 4.4. 在这样的一个分组频数分布中，用来表示各组的符号(例如，128－130)称为**分组区间**. 表 4.4 包含六列：分组区间(这里表示为时间)、**组限**、**组界**、**组中值**、**计数**和**频数**.

表 4.4

时间 (min)	组 限 下限—上限	组 界 下界—上界	组中值 m_i(min)	计 数	频数 f_i
128—130	128—130	127.5—130.5	129	|||	3
131—133	131—133	130.5—133.5	132	|	1
134—136	134—136	133.5—136.5	135	||||	4
137—139	137—139	136.5—139.5	138	|||	3
140—142	140—142	139.5—142.5	141	卌 ||	7
143—145	143—145	142.5—145.5	144	卌 卌 ||	12
Σ					30

组限是指各组取值范围的最大测量值(上限)和最小测量值(下限). 它们定义了分组区间.

组界(也称为**真实组限**)表示**蕴含组范围**. 由 2.10 节可知,一个近似测量值的最后一位数字实际上可以看成一个区间,即蕴含范围. 因此,128 分钟实际上是 127.5- 128.5 分钟之间的某个值,130 分钟实际上是 129.5 --130.5 分钟之间的某个值. 这两个值定义了第 1 组的分组区间(128—130),所以这个组的下组界是 127.5 分钟,上组界是 130.5 分钟. 类似的,第 2 组(131—133)的下组界是 130.5 分钟(也是第 1 组的上组界),上组界是 133.5 分钟,….第 6 组(143—145)的组界是 142.5 分钟和 145.5 分钟.

组中值是恰好处于组中间位置的值(中点),也称为**组标**,通常可以用组的上、下限之和除以 2 得到,在这里用 m_i 表示,其中 i 表示第 i 组. 因此,第一组的组中值是

$$m_1 = \frac{128+130}{2} = 129$$

计数列和频数列与未分组频数分布中的相同,只是在这里它们表示各组的观测值个数,而不是各类的观测值个数.

组距是指各组取值范围的大小,可以用相继组的下限之差或上限之差计算得到. 例如,用相继组的下限计算,第一组的组距是:131－128＝3. 事实上,这个表中的所有组距都是 3,即这个表是等组距的. 对于等组距频数分布,也可以由相继组中值相减得到组距. 这里,用前两个组中值 129,132 计算得到所有组的组距都是 3.

4.5 分组频数分布和分组百分数分布

对于分组分布,各组的相对频数和百分数的计算与未分组分布相同.

例 4.6 将表 4.4 中的组界、组中值和计数这三列去掉,增加表示相对频数和百分数的列.

解 ☞ 所要求的汇总表见表 4.5.

表 4.5

时间 (min)	组限 下限—上限	频数 f_i	相对频数 f_i/N	百分数 $[(f_i/N)\times 100]\%$
128—130	128—130	3	0.1000	10.00
131—133	131—133	1	0.0333	3.33
134—136	134—136	4	0.1333	13.33
137—139	137—139	3	0.1000	10.00
140—142	140—142	7	0.2333	23.33
143—145	143—145	12	0.4000	40.00
Σ		30		

4.6　未分组分布转换成分组分布时应遵循的原则

将未分组分布转换成分组分布时,关键是使数据在分组后其最显著的趋势能很明显.由于分组时涉及到个人的观察力和判断力,所以这个问题没有唯一的最好解决方案.但是,统计学给出了这样一些原则:

(1) 组数不小于5,且不大于20;

(2) 尽可能对所有组使用相同的组距;

(3) 组距可以是偶数也可以是奇数,但是一般建议用奇数,以使组中值仍落在测量尺度上;

(4) 组距可以是任意的,但建议组距取为5,10,50,100,500等数的倍数;

(5) 若组距相等,应确保:极差<组数×组距;

(6) 最小组的下限的选取应使最小测量值包含在该组中,最大组的上限的选取应使最大测量值包含在该组中;

(7) 为避免夸大比 x_s 小的类及比 x_l 大的类,最小组的下限应尽量靠近最小测量值,最大组的上限应尽量靠近最大测量值;

(8) 通常,数据越多,分的组数应越多.

例 4.7　运用上述转换原则,说明如何将表4.3中的频数分布转换成表4.4中分组频数分布.

解 ☞　为了运用上述转换原则,首先计算这组数据的极差

$$极差 = x_l - x_s = 145\ \text{min} - 129\ \text{min} = 16\ \text{min}$$

接下来应同时考虑下列因素:

(a) 希望各组的组距相等(原则2),且是一个奇数(原则3),并尽可能是5的倍数(原则4);

(b) 至少要分成5组(原则1),并且使16<组数×组距(原则5);

(c) 最小组的下限的选取应使129 min包含在该组中,最大组的上限的选取应使145 min包含在该组中(原则6);

(d) 最小组的下限应尽量靠近129 min,最大组的上限应尽量靠近145 min(原则7);

(e) 使数据中的重要趋势变得最明显,这是一个**主观因素**.

只有下列四种组数和组距的组合能满足上述大部分条件:

	组数		组距
16<[20=	5	×	4]
16<[18=	6	×	3]
16<[21=	7	×	3]
16<[18=	9	×	2]

通过对这几种组合分组分布的研究,我们认为最佳选择是组数为6,组距为3,也就是表4.4中使用的分组分布.这是因为:它的组距是个奇数,组中值恰好是尺度单位(例129,132),最小组的下限(128 min)低于且非常靠近最小测量值129 min,最大组的上限则正好是最大测量值145 min,另外,该分组频数分布有一个明显的特征,即频数随着时间的增加而增加.

4.7　开端点组分布和不等组距

只有一个组限,上组限或下组限的组称为**开端点组**.至少包含一个开端点组的分组分布称为**开端点组分布**.

如果测量值中有少数一些非常大或非常小的值,这些值远离大部分数据集中的位置,这时可以用开端点组.使用开端点组的另一个原因是对信息的保密.例如,公布统计专业班级的第二门考试成绩(见附录中表A.2的第3列)时,为了避免使成绩很差的学生感到不好意思,可以使用没有下限的开端点分组分布,如开端点组为小于等于64.

开端点组分布中常具有多种组距.采用不等组距是为了强调具有重大政治或经济意义的某些组,或者是为了将包含极少测量值的区域合并.

由于开端点组的一些性质(例如组限、组界、组距和组中值)无定义,使得数据的图形化(见第五章)较困难,且不能由数据计算大部分描述统计量(见第六章和第七章),所以如果可能一般应避免使用开端点组.

例 4.8 表 4.6 中的分组频数分布摘自《美国统计摘要》(1995 年)的第 74 页. 表中将 1992 年美国的新生婴儿按照母亲当年的年龄分为 6 个组,新生婴儿的个数以 1000 为单位. 利用给出的信息,添加以下各列:组限、组界、组距和组中值.

表 4.6

母亲年龄	1992 (新生婴儿数(千))
20 岁以下	518
20—24 岁	1,070
25—29 岁	1,179
30—34 岁	895
35—39 岁	345
40 岁及以上	58

解 ☞ 表 4.6 中的分布举例说明了开端点组及开端点组频数分布. "20 岁以下"组是一个开端点组,因为只有上限 19 岁,而没有下限. 类似的"40 岁及以上"组也是开端点组,因为只有下限 40 岁,但没有上限. 表 4.7 为添加各所需列后的频数分布表. 表中,问号表示由于开端点组的性质而无法确定的数值.

表 4.7

母亲年龄	组限	组界	组距	组中值 m_i(年)	1992(新生婴儿数(千)) f_i
20 岁以下	?—19	?—19.5	?	?	518
20—24 岁	20—24	19.5—24.5	5	22	1,070
25—29 岁	25—29	24.5—29.5	5	27	1,179
30—34 岁	30—34	29.5—34.5	5	32	895
35—39 岁	35—39	34.5—39.5	5	37	345
40 岁及以上	40—?	39.5—?	?	?	58
$\sum$					4,065

例 4.9 表 4.8 中的开端点组频数分布摘自《美国统计简要》(1992 年,《美国统计摘要》的补充). 这是 1991 年美国居民按年龄分组的频数分布. 利用给出的信息,添加以下各列:组限、组界、组距和组中值.

表 4.8

总体	1991 (人数(百万))
5 岁以下	19.2
5—17 岁	45.9
18—24 岁	26.4
25—34 岁	42.9
35—44 岁	39.3
45—64 岁	46.7
65 岁及以上	31.8

解 ☞ 表 4.9 为添加各所需列后的频数分布表. 表中,问号同样是表示由于开端点组的性质而无法确定的数值. 这个分布举例说明了开端点组频数分布能用不等组距来构造. 这里,有两个开端点组,四个不同的组距。

表 4.9

总体	组限	组界	组距	组中值 m_i(年)	1991(新生婴儿数(千)) f_i
5 岁以下	?—4	?—4.5	?	?	19.2
5—17 岁	5—17	4.5—17.5	13	11	45.9
18—24 岁	18—24	17.5—24.5	7	21	26.4
25—34 岁	25—34	24.5—34.5	10	29.5	42.9
35—44 岁	35—44	34.5—44.5	10	39.5	39.3
45—64 岁	45—64	44.5—64.5	20	54.5	46.7
65 岁及以上	65—?	64.5—?	?	?	31.8
$\sum$					252.2

4.8　"小于式"累积分布

"小于式"累积频数分布给出了数据集中小于给定值的数据个数. 为了从未分组频数分布中得到这样一个分布,需要将各类的频数进行**累加**.

如果样本数据由连续的因此是近似的测量值组成,则"小于式"累积是指累积到某个测量类的蕴含范围(见 2.10 节)的上界为止. 相应的累积频数就是样本中小于该边界值的所有测量值的个数.

如果样本数据是离散比例,则由于样本值为准确的测量值(见 2.10 节),它没有蕴含范围,所以它的"小于式"累积频数的计算方法不同. 其"小于式"累积是指累积到类本身值为止,相应的累积频数就是样本中小于该值的所有测量值的个数.

例 4.10　将表 4.1 中的频数分布转换成"小于式"累积频数分布.

解 ☞　表 4.10 给出了转换后的"小于式"累积频数分布,称为**"小于式"累积频数表**. 表 4.1 中的最上端的值为 1.0 kg(x_s),是一个近似测量值,它的蕴含范围为 0.95 kg 到 1.05 kg. 因此,表 4.10 中的累积从 0.95 kg 开始,并指出没有数小于 0.95 kg. 由于 1.0 kg 中有两个数,它们位于 0.95 kg 到 1.05 kg 之间的某处,所以累积指出有两个数小于 1.05 kg. 由于 1.1 kg 中(1.05 kg 到 1.15 kg)没有数,所以累积指出有两个数小于 1.15 kg. 在下一个类,1.2 kg 中(1.15 kg 到 1.25 kg)有四个数值,所以有六个数小于 1.25 kg. 继续这个累积过程,直到最后一个类,1.5 kg(x_l,1.45 kg 到 1.55 kg),累积指出样本中所有 20 个数值都小于 1.5 kg.

表 4.10

重量(kg)	累积频数
小于 0.95	0
小于 1.05	2
小于 1.15	2
小于 1.25	6
小于 1.35	14
小于 1.45	18
小于 1.55	20

4.9　"大于等于式"累积分布

"大于等于式"累积频数分布给出了数据集中等于或大于给定值的数据个数. 对于近似测量值,累积频数给出了从 x_l 到 x_s 等于或大于某类的蕴含范围的下界的数据个数. 对于精确测量值,累积频数给出了等于或大于某给定类值的数据个数.

"小于式"和"大于等于式"分布是累积频数分布中最常见的形式,但是,有时也会构造**"小于等于式"累积频数分布**(对于近似测量值,累积频数给出了等于或小于某个类的蕴含范围的上界的数据个数. 对于精确测量值,累积频数给出了等于或小于类本身值的数据个数),或**"大于式"累积频数分布**(对于近似测量值,累积频数给出了大于某类的蕴含范围的下界的数据个数. 对于精确测量值,累积频数给出了大于某类本身值的数据个数).

例 4.11　将表 4.1 中的频数分布转换成"大于等于式"累积频数分布.

解 ☞　表 4.11 给出了转换后的"大于等于式"分布,称为**"大于等于式"累积频数分布表**. 为了得到这个分布表,需从最大观测值 x_l 的下一个类的下界开始累积. 由于最大观测值 x_l 等于 1.5 kg,所以其下一个类(1.6 kg)的下界是 1.55 kg. 因此,表 4.11中的累积从 1.55 kg 开始,并指出没有数大于等于 1.55 kg. 由于 x_l 中有两个数,位于 1.45 kg 到 1.55 kg 之间,所以累积指出有两个数大于等于 1.45 kg. 由于下一个类,1.4 kg 中(1.35 kg 到 1.45 kg)中有四个数值,所以累积指出有六个数大于等于 1.35 kg. 继续这个累积过程,直到 x_s,1.0 kg(0.95 kg 到 1.05 kg),累积指出样本中所有 20 个数值都大于等于 0.95 kg.

表 4.11

重量(kg)	累积频数
大于等于 0.95	20
大于等于 1.05	18
大于等于 1.15	18
大于等于 1.25	14
大于等于 1.35	6
大于等于 1.45	2
大于等于 1.55	0

4.10 分组累积分布

分组"小于式"累积分布通常累积到本组的上一组的上界，但也可以累积到本组的下限.类似的，**分组"大于等于式"累积分布**可以累积到本组的下一组的下界或本组的上限.

例 4.12 将表 4.4 中马拉松时间的分组频数分布转换成分组"小于式"累积频数分布，要求用组的上界进行累积.

解 表 4.12 给出了转换后的分组"小于式"累积频数分布.由表中可以看出，没有数小于 127.5 min(比包含 x_i 的组更小的上一个组的上界)，有三个数小于 130.5 min(包含 x_i 的组的上界)，等等，直到所有 30 个数都小于 145.5 min(包含 x_i 的组的上界).

表 4.12

时间(min)	累积频数
小于 127.5	0
小于 130.5	3
小于 133.5	4
小于 136.5	8
小于 139.5	11
小于 142.5	18
小于 145.5	30

习题解答

阵列和极差

4.1 附表 A.2 中第 5 列是 16 个女生的体重测量值，而第 9 列中用 * 标志的是例 3.5 中得到的 16 名学生(男生和女生)的简单随机样本的体重测量值，请分别计算这两组数据的极差，并指出哪个极差更大，为什么?

解 对于 16 个女生的体重测量值，

$$极差=136\ \text{lb}-105\ \text{lb}=31\ \text{lb}$$

对于简单随机样本的体重测量值，

$$极差=186\ \text{lb}-115\ \text{lb}=71\ \text{lb}$$

简单随机样本的极差比女生的两倍还多，这是因为男生一般都比女生更重，而这个简单随机样本正好包括了四个最轻女生中的一个和四个最重男生中的一个.

4.2 对于习题 3.12 中得到的比例分层随机样本(附表 A.2 中第 10 列中用 * 标志的学生)和习题 3.24 中得到的系统随机样本(附表 A.2 中第 11 列中用 * 标志的学生)，请分别将他们的学期论文成绩(第 8 列)排成升序阵列，并计算极差.

解 学期论文成绩在这里是用字母表示的(从 A 表示优秀到 B,C,D，最后 F 表示不及格)，是次序测量尺度(见 2.5 节).这种测量值可以进行排序，但是无法用加减法确定测量值之间的精确距离，所以对于这些次序测量值，我们只能确定它们的升序阵列，而无法计算极差.

由表 A.2 可知，比例分层随机样本的成绩为：$F,B,C,A,B,B,A,B,C,A,A,A,B,B,B,A$(按照它们在附表 A.2 中的出现顺序排列).

其升序阵列为：$F,C,C,B,B,B,B,B,B,B,A,A,A,A,A,A$.

由表 A.2 可知，系统随机样本的成绩为：$B,B,A,A,B,C,B,C,D,B,C,A,B,A,F,B$(按照它们在附表 A.2 中的出现顺序排列).

其升序阵列为：$F,D,C,C,C,B,B,B,B,B,B,B,A,A,A,A$.

注：虽然对于次序水平的数据计算其极差是不合适的，但是这里可以在非统计意义下应用极差.在日常用语中，极差是指一组事物的范围，因此可以说"所有样本的成绩范围都是从 F 到 A.

4.3 对于习题 3.13 中得到的单阶段整群随机样本(附表 A.2 中第 12 列中用 * 标志的学生)，请将他们的期末考试成绩(第 3 列)排成降序阵列，并计算极差.

解 在这里，成绩是用百分制表示的，是离散比例尺度测量(见 2.8 节)，所以阵列和极差都是可以得到的.按照附表 A.2 中的顺序，单阶段整群样本的成绩为：88，78，97，82，90，94，79，90，74，88，86，81，85，64，64，92.其降序阵列为：97，94，92，90，90，88，88，86，85，82，81，79，78，74，64，64.极差为：97－64＝33.

4.4 为什么附表 A.2 中女生的身高精确到 1/4 英寸,而不是精确到英寸?

解 ☞ 为了使数据集中的测量值足够分离,所使用的测量尺度应在 30 到 300 单位步长之间. 如果使用英寸为单位步长,那么这些女生使用的测量尺度是 60 英寸到 69 英寸,这时只包括了 10 个类. 如果测量值精确到 1/4 英寸,即使用 1/4 英寸为单位步长,那么这些女生使用的测量尺度的类增加到 37 个,介于 30 到 300 之间.

分布:频数、相对频数和百分数

4.5 对于表 4.2,计算(a) $\sum_{i=1}^{k}(f_i/n)$, (b) $\sum_{i=1}^{k}[(f_i/n)\times 100]\%$

解 ☞ (a) 在舍入误差允许范围内,相对频数列的和总是 1.0,所以

$$\sum_{i=1}^{k}(f_i/n)=1.0$$

(b) 在舍入误差允许范围内,百分数列的和总是 100

$$\sum_{i=1}^{k}[(f_i/n)\times 100]\%=100\%$$

4.6 研究一种从中国进口的猪的特征,这种类型的母猪每窝的产仔数量要比一般的美国本土猪要多. 一般的美国本土猪每窝的产仔数量为 10 到 12 只,而这种类型的母猪每窝的产仔数量为 16 到 20 只. 表 4.13 是这种母猪的 50 窝产仔数量的频数(相对频数)分布表,请填空完成该表.

解 ☞ 在相对频数这一列中有一个缺失值,即一窝产仔数量为 21 只的相对频数. 由于相对频数列的和总是 1.00[见习题 4.5(a)],而不包括 21 只的所有相对频数之和为 0.92,所以可计算得到一窝产仔数量为 21 只的相对频数是 1.00−0.92=0.08.

由于相对频数=频数/n,而 n=50,所以在这里,频数=相对频数×50.

表 4.14 是完成后的分布表.

表 4.13

产仔数量 x_i	频数 f_i	相对频数 f_i/n
15		0.02
16		0.04
17		0.10
18		0.40
19		0.24
20		0.12
21		

表 4.14

产仔数量 x_i	频数 f_i	相对频数 f_i/n
15	1	0.02
16	2	0.04
17	5	0.10
18	20	0.40
19	12	0.24
20	6	0.12
21	4	0.08
$\sum$	50	1.00

4.7 将附表 A.2 中的学期论文成绩(第 8 列)分成男生和女生的成绩表,每个表中包括以下各列:成绩、频数、相对频数和百分数.

解 ☞ 女生的统计表见表 4.15,男生的统计表见表 4.16.

表 4.15

成绩 x_i	频数 f_i	相对频数 f_i/n	百分数 $[(f_i/n)\times 100]\%$
F	1	0.0625	6.25
D	1	0.0625	6.25
C	3	0.1875	18.75
B	6	0.3750	37.50
A	5	0.3125	31.25
$\sum$	16	1.0000	100.00%

表 4.16

成绩 x_i	频数 f_i	相对频数 f_i/n	百分数 $[(f_i/n)\times 100]\%$
F	2	0.0417	4.17
D	4	0.0833	8.33
C	10	0.2083	20.83
B	20	0.4167	41.67
A	12	0.2500	25.00
$\sum$	48	1.0000	100.00%

4.8 将附表 A.2 中的头发颜色(第 7 列)分成男生和女生的头发颜色表,每个表中包括以下各列:头发颜色、频数、相对频数和百分数.

解 ☞ 女生的统计表见表 4.17,男生的统计表见表 4.18.

表 4.17

头发颜色 x_i	频数 f_i	相对频数 f_i/n	百分数 $[(f_i/n)\times100]\%$
黑色	3	0.1875	18.75
金黄色	8	0.5000	50.00
棕色	4	0.2500	25.00
红色	1	0.0625	6.25
$\sum$	16	1.0000	100.00%

表 4.18

头发颜色 x_i	频数 f_i	相对频数 f_i/n	百分数 $[(f_i/n)\times100]\%$
黑色	10	0.2083	20.83
金黄色	14	0.2917	29.17
棕色	20	0.1167	11.67
红色	4	0.0833	8.33
$\sum$	48	1.0000	100.00%

分组分布

4.9 请问对于同一数据集,何时选用分组分布而不是未分组分布?

解 ☞ 任何分布函数,分组分布或未分组分布,都是用来表示定量信息的汇总.所以,要决定选择何种形式的分布,其目的就是要提供最简明、最易于理解的汇总——最能清楚表示重点的汇总.为了达到这个目的,通常我们选用分组分布而不是未分组分布.

未分组分布必须包括测量尺度中使用部分的所有单位步长,例如习题 4.4 中推荐的 30 到 300 之间的单位步长,这么多的单位步长使得数据分布特别分散,并且有很多空的或低频数的类.分组频数分布将这些类压缩成一些较少的组,更能清楚地表示数据的重要信息.例如,表 4.4 中的分组频数分布将原始的频数分布(表 4.3)中的 17 个类压缩成了 6 组,去掉了很多空的和低频数的类.这使得数据中的重要趋势变得更明显:频数随着时间的增加迅速增加.

4.10 由题中给出信息,完成表 4.19 中空白部分.

表 4.19

长度(mm)	组限	组界	组中值 m_i(mm)	组距	频数 f_i	相对频数 f_i/n
1.1—1.2					5	
1.3—1.4					6	
1.5—1.6					6	
1.7—1.8					8	
1.9—2.0					25	
$\sum$					50	

解 ☞ 表 4.20 为完成后的表.以表中最上面一行为例:组限为 1.1—1.2,与分组区间相同[表中标为长度(mm)];组下界是 1.05,是近似测量值 1.1mm 的隐含范围的下界,组上界是 1.25,是近似测量值 1.2mm 的隐含范围的上界;组中值为 $m_1=(1.1+1.2)/2=1.15$;组距为 $1.3-1.1=0.2$;这个组的相对频数为 $f_1/n=5/50=0.10$.

表 4.20

长度(mm)	组限	组界	组中值 m_i(mm)	组距	频数 f_i	相对频数 f_i/n
1.1—1.2	1.1—1.2	1.05—1.25	1.15	0.2	5	0.10
1.3—1.4	1.3—1.4	1.25—1.45	1.35	0.2	6	0.12
1.5—1.6	1.5—1.6	1.45—1.65	1.55	0.2	6	0.12
1.7—1.8	1.7—1.8	1.65—1.85	1.75	0.2	8	0.16
1.9—2.0	1.9—2.0	1.85—2.05	1.95	0.2	25	0.50
$\sum$					50	1.00

4.11 在例 4.5 中,我们给出了波士顿马拉松比赛的前 30 名男运动员的成绩.现在我们将这一数据集扩展为比赛中的前 90 名男运动员的成绩,以升序阵列表示:129,130,130,133,134,135,136,136,138,138,138,141,141,141,142,142,142,142,143,143,143,143,143,144,144,145,145,145,145,145,146,146,146,146,147,147,147,148,148,148,148,148,148,149,149,149,149,149,150,150,150,150,151,151,151,151,152,152,152,152,152,152,152,152,152,153,153,153,153,153,153,153,153,153,153,153,154,154,154,154,154,154,154,154,154,155,155,155,155,155.

应用 4.6 节给出的原则,将前 85 个运动员的成绩整理成分组分布,要求分成 9 组,分布表中应包括分组区间、组界、组中值、频数和百分数各列.

解 ☞　前 85 个运动员的成绩的极差=154 min−129 min=25 min.我们希望频数分布表中各组的组距相等,且是一个奇数,还要满足 25<9×组距,所以最优组距仍然是 3.表 4.21 为完成后的分布表.

表 4.21

时间(秒)	组界	组中值 m_i(秒)	频数 f_i	百分数 $[(f_i/n)\times100]\%$
128—130	127.5—130.5	129	3	3.53
131—133	130.5—133.5	132	1	1.18
134—136	133.5—136.5	135	4	4.71
137—139	136.5—139.5	138	3	3.53
140—142	139.5—142.5	141	7	8.24
143—145	142.5—145.5	144	12	14.12
146—148	145.5—148.5	147	13	15.29
149—151	148.5—151.5	150	13	15.29
152—154	151.5—154.5	153	29	34.12
$\sum$			85	100.01%

4.12 应用 4.6 节中给出的原则,将附表 A.2 中 64 个学生的期末考试成绩(第 3 列)整理成分组分布,要求组距为 5 的倍数,分布表中应包括分组区间、组界、组中值、计数、频数和百分数各列.

解 ☞　根据附表 A.2 可计算得到极差=99−49=50.由于要求组距是 5 的倍数,所以选择组距为 5.由 4.6 节中的原则知,组数应满足:50<组数×5,所以最优组数是 11.在这 64 个学生的成绩中,最高分是 99 分,所以最大组的组限有两种选择:95—99 或 96—100.为了强调全班没有一个学生得到满分 100 分,所以我们选择 95—99.这个选择决定了分布中有 11 个组区间,表 4.22 为完成后的分布表.由于数据是离散比例水平的,所以组区间、组限和组界都相同.每一组的组中值是其组限之和除以 2.附表 A.2 中每一个学生成绩都在计数列中相应的类中用一个计数标记给出,最后将计数转换成频数和百分数.

表 4.22

第二次考试	组界	组中值 m_i	计数	频数 f_i	百分数 $[(f_i/n)\times100]\%$
45—49	45—49	47	|	1	1.5625
50—54	50—54	52		0	0.0000
55—59	55—59	57	||||	4	6.2500
60—64	60—64	62	|||	3	4.6875
65—69	65—69	67	卌	5	7.8125
70—74	70—74	72	|||	3	4.6875
75—79	75—79	77	卌 |	6	9.3750
80—84	80—84	82	卌 卌 |	11	17.1875
85—89	85—89	87	卌 ||||	9	14.0625
90—94	90—94	92	卌 卌 卌 ||	17	26.5625
95—99	95—99	97	卌	5	7.8125
$\sum$				64	100.0000%

4.13　应用 4.6 节中给出的原则，将附表 A.2 中 64 个学生的体重(第 5 列)整理成分组分布，分布表中应包括分组区间、组中值、计数、频数各列，并且要求将各组的计数和频数分别按照女生、男生和总计三种情况列出.

解 ☞　根据附表 A.2 可计算得到 64 个体重的极差＝194 lb－105 lb＝89 lb. 对于这个极差，最优组数为 9，最优组距为 10. 表 4.23 为完成后的分布表.

表 4.23

体重(lb)	组中值 m_i(lb)	女生		男生		总数	
		计数	频数 f_i	计数	频数 f_i	计数	频数 f_i
105—114	109.5	|||	3		0	|||	3
115—124	119.5	卌 |||	8	|||	3	卌 卌 |	11
125—134	129.5	||||	4	|	1	卌	5
135—144	139.5	|	1	卌 |	6	卌 ||	7
145—154	149.5		0	卌 卌 |	11	卌 卌 |	11
155—164	159.5		0	卌 卌 |	11	卌 卌 |	11
165—174	169.5		0	卌 |||	8	卌 |||	8
175—184	179.5		0	||||	4	||||	4
185—194	189.5		0	||||	4	||||	4
$\sum$			16		48		64

4.14　名义型和次序型数据可以整理成分组分布吗？

解 ☞　不可以. 因为分组分布是仅限于间隔测量和比例测量数据的分布. 正如 4.4 节中所定义的，分组分布是将原始测量尺度分成若干个有序的组，并相应明确定义了非主观的组限、组界、组距和组中值. 名义型和次序型数据是不可能进行如此分组的.

但是，为了表示或分析的目的，经常重新组织以形成更大更综合的类来表示和分析名义水平和次序水平的数据. 例如，学期论文成绩的频数分布(见表 4.15 和表 4.16)可以重新组织成下面的类：(A)、(B)，和(C 或以下)，以强调好成绩和差成绩；或者为了强调一种头发颜色，将头发颜色的分布(见表 4.17 和表 4.18)重新组织成金色和非金色. 但是，名义水平和次序水平分布的重新组织不称为分组分布.

开端点组分布和不等组距

4.15　表 4.6 是一个开端点组分布，请问这个分布的构造是否遵循了 4.6 节提出的原则？

解 ☞　这个分布的构造基本上遵循了 4.6 节提出的原则. 第 1 个原则建议使用 5—20 个组，表 4.6 中包含了 6 组. 第 2 个原则是“尽可能对所有组使用相同的组距”，在这个分布中，除了开端点组的组距无法知道，其它 4 个组的组距都是 5. 这个组距是奇数并且是 5 的倍数，这与第 3 个和第 4 个原则一致. 对于开端点组，没有用到第 5 个和第 7 个原则，但是可以假定遵循了第 6 个原则.

4.16　表 4.24 中的开端点组频数分布摘自《美国统计摘要》(1995 年)的第 471 页. 表中将 1993 年美国家庭按照家庭收入分为 9 个组，每组中家庭的个数以 1000 为单位. 利用给出的信息，增加以下各列：组界、组距和百分数.

表 4.24

家庭收入	家庭个数(以 1000 为单位)
\$ 5,000 以下	4,407
\$ 5,000—\$ 9,999	9,467
\$ 10,000—\$ 14,999	8,956
\$ 15,000—\$ 19,999	8,319
\$ 20,000—\$ 24,999	8,103
\$ 25,000—\$ 34,999	14,318
\$ 35,000—\$ 49,999	15,791
\$ 50,000—\$ 74,999	13,632
\$ 75,000 以上	12,114

解 ☞　表 4.25 为增加各所需列后的频数分布表. 表中问号表示由于开端点组的性质而无法确定的数值. 尽管家庭收入是离散比例水平(见 2.8 节)，但是此处组界不等于组限. 在美国，收入不能再细分至比美分更深的水平，所以收入是离散比例水平. 但是此处的收入是精确到美元，所以组界和组限不相同. 因此，可以用相邻的组限之间计算的中点作为组界.

表 4.25

家庭收入	组界	组距	家庭个数 (以 1000 为单位) f_i	百分数 $[(f_i/n)\times 100]\%$
\$5,000 以下	? - 4,999.5	?	4,407	4.5383
\$5,000—\$9,999	4,999.5—9,999.5	5,000	9,467	9.7490
\$10,000—\$14,999	9,999.5—14,999.5	5,000	8,956	9.2228
\$15,000—\$19,999	14,999.5—19,999.5	5,000	8,319	8.5668
\$20,000—\$24,999	19,999.5—24,999.5	5,000	8,103	8.3444
\$25,000—\$34,999	24,999.5—34,999.5	10,000	14,318	14.7446
\$35,000—\$49,999	34,999.5—49,999.5	15,000	15,791	16.2614
\$50,000—\$74,999	49,999.5—74,999.5	25,000	15,632	16.0977
\$75,000 以上	74,999.5—?	?	12,114	12.4749
$\sum$			97,107	100.0000%

4.17 下列降序阵列是某高尔夫球联赛中 70 名选手赢得的奖金(以美元为单位),括号中是赢得该奖金额度的选手人数:

200,000(1);83,333(3);45,000(1);40,000(1);36,250(2);30,000(3);21,900(5);15,000(7);10,000(3);7,535(7);5,750(7);4,260(5);3,220(5);2,750(2);2,490(5);2,380(3);2,330(2);2,290(2);2,260(1);2,240(1);2,220(2);2,180(1);2,170(1)

某报纸的体育新闻记者在作这次联赛的专访,他想使用一个分组频数分布说明如下观点:这次联赛中只有少数选手赢得了巨额奖金,绝大部分选手赢得的奖金都非常少.将奖金精确到 1,000 美元后,分成 10 组,没有开端点组,其中第一组为 \$2,000—\$4,000.要说明这一观点,只需使用分组区间和频数两列.请给出满足上述条件的一个频数分布表,要求列出分组区间、组界、组距、计数和频数.

解 ☞　表 4.26 为满足条件的一个频数分布表.表中有 10 组,并且没有开端点组.为了说明随着奖金额度的递增获奖人数急剧下降,该分布使用了 7 个不同的组距.

表 4.26

奖金(\$)	组界	组距	计数	频数 f_i
2,000—4,000	1,500—4,500	3,000	卌 卌 卌 卌 卌 卌	30
5,000—10,000	4,500—10,500	6,000	卌 卌 卌 |	17
11,000—20,000	10,500—20,500	10,000	卌 ||	7
21,000—30,000	20,500—30,500	10,000	卌 |||	8
31,000—40,000	30,500—40,500	10,000	||	3
41,000—45,000	40,500—45,500	5,000	|	1
46,000—81,000	45,500—81,500	36,000		0
82,000—83,000	81,500—83,500	2,000	||	3
84,000—198,000	83,500—198,500	115,000		0
199,000—200,000	198,500—200,500	2,000	|	1
$\sum$				70

累积分布

4.18 对于表 4.10 中的"小于式"累积频数分布,添加表示"小于式"累积相对频数分布和"小于式"累积百分数分布的列.

解 ☞　表 4.27 为添加各所需列后的分布表.考虑只有近似测量值的未分组频数分布,其"小于式"累积相对频数分布用来说明数据集中所有小于某个测量类的蕴含范围上界的数据所占的比例.而"小于式"累积百分数分布则说明小于该上界的数据所占的百分数.

表 4.27

重量(kg)	累积频数	累积相对频数	累积百分数
小于 0.95	0	0.0	0
小于 1.05	2	0.1	10
小于 1.15	2	0.1	10
小于 1.25	6	0.3	30
小于 1.35	14	0.7	70
小于 1.45	18	0.9	90
小于 1.55	20	1.0	100

为了将表 4.10 中的"小于式"累积频数分布转换成"小于式"累积相对频数分布,只需将每个累积频数除以样本容量($n=20$),例如 0/20=0.0,2/20=0.1 等等.然后,将这些相对频数的值转换成百分数,只需将每个"小于式"累积相对频数值都乘以 100,就得到了"小于式"累积百分数分布,例如 0.0×100=0.0,1×100=10.

4.19 将表 4.14 中的产仔数量的频数分布和相对频数分布转换成"小于式"累积频数和累积相对频数分布.

表 4.28

产仔数量	累积频数	累积相对频数
小于 15	0	0.00
小于 16	1	0.02
小于 17	3	0.06
小于 18	8	0.16
小于 19	28	0.56
小于 20	40	0.80
小于 21	46	0.92
小于 22	50	1.00

解 ☞ 产仔数量是准确测量值(见 2.10 节),所以,其"小于式"累积是从 x_1 到 x_l 按类累积,相应的累积频数就是样本中小于该类本身值的所有测量值的个数.

为了得到"小于式"累积相对频数分布,需要将每个累积频数除以样本容量.利用该方法得到这个问题所要求的累积分布,见表 4.28.由表 4.14 有,$x_1=15$ 的频数是 1,因此表 4.28 的最上面一行指出没有数小于 15,小于 15 的数的比例是 0/50=0.0.在下一行中,累积指出有 1 个数小于 16(比例=1/50=0.02).继续这个过程直到 $x_l=21$,有 46 个数小于 21(比例=46/50=0.92).为了完成累积,将 $x_l=21$ 的下一个类包括进来,并且认为有 50 个数小于 22(比例=50/50=1.00).

4.20 表 4.11 是一个"大于等于式"累积频数分布,请添加表明"大于等于式"累积相对频数分布和"大于等于式"累积百分数分布的列.

解 ☞ 表 4.29 为增加各所要求列后的分布.根据习题 4.18 中频数到相对频数的转换,要将表4.11 中的"大于等于式"累积频数分布转换成"大于等于式"累积相对频数分布,只需将累积频数除以样本容量($n=20$).由于有 20 个数大于等于 0.95 kg,所以大于等于 0.95 kg 的数的比例是 20/20=1.0.在另一个极端,有 0 个数大于等于 1.55 kg,所以大于等于 1.55 kg 的数的比例是 0/20=0.0.要将"大于等于式"累积相对频数分布转换成"大于等于式"累积百分数分布,只需将累积相对频数乘以 100.于是,1.0×100=100,0.9×100=90,等等,0.0×100=0.

表 4.29

重量(kg)	累积频数	累积相对频数	累积百分数
大于等于 0.95	20	1.0	100
大于等于 1.05	18	0.9	90
大于等于 1.15	18	0.9	90
大于等于 1.25	14	0.7	70
大于等于 1.35	6	0.3	30
大于等于 1.45	2	0.1	10
大于等于 1.55	0	0.0	0

4.21 将表 4.22 学生期末考试成绩的分组频数分布转换成分组"大于等于式"累积频数和累积百分数分布.

解 ☞ 表 4.30 即为所要求的分组累积分布.在这个问题中,学生的成绩是精确的离散比例型测量值(见 2.8 节),所以组界和组限相等.此处无论考虑哪一个,都有 100%的成绩大于等于 45,98.4%的

成绩大于等于 50，等等，直到 0.0%的成绩等于 100.

表 4.30

第二次考试	累积频数	累积百分数
大于等于 45	64	100.0
大于等于 50	63	98.4
大于等于 55	63	98.4
大于等于 60	59	92.2
大于等于 65	56	87.5
大于等于 70	51	79.7
大于等于 75	48	75.0
大于等于 80	42	65.6
大于等于 85	31	48.4
大于等于 90	22	34.4
大于等于 95	5	7.8
大于等于 100	0	0.0

表 4.31

奖金($)	累积百分数
小于 1,500	0.6
小于 4,500	42.9
小于 10,500	67.1
小于 20,500	77.1
小于 30,500	88.6
小于 40,500	92.9
小于 45,500	94.3
小于 81,500	94.3
小于 83,500	98.6
小于 198,500	98.6
小于 200,500	100.0

4.22 将表 4.26 中高尔夫球选手所获奖金的分组频数分布转换成分组"小于式"累积百分数分布，要求在累积中使用组上界.

解 ☞　表 4.31 即为所要求的分组累积分布.

注：这是一个不等组距累积分布的例子，其构造方法与等组距累积分布的相同.

4.23 对 50 个想加入警察队伍的人进行反应时间测试. 测试结果(精确到十分之一秒)汇总成"大于等于式"累积百分数分布，如表 4.32 所示. 累积使用的是类的蕴含范围下界. 如果警察局需要的人是在这次测试中反应时间小于等于 0.5 秒的人，请问这 50 人中有多少人通过这一测试？有多少人比所要求时间慢 0.1 秒？

表 4.32

反应时间(sec)	累积百分数
大于等于 0.15	100
大于等于 0.25	98
大于等于 0.35	94
大于等于 0.45	86
大于等于 0.55	76
大于等于 0.65	50
大于等于 0.75	30
大于等于 0.85	18
大于等于 0.95	8
大于等于 1.05	0

解 ☞　从表 4.32 可以看出，反应时间大于等于 0.55 秒的人占 76%，所以反应时间小于等于 0.5 秒的人占 24%. 也就是说，有 $0.24\times50=12$ 的人通过了测试.

由于有 76%的人比要求的时间慢 0.1 秒或更多，而有 50%的人比要求的时间慢 0.2 秒或更多，所以有 26%的人，即 13 个人比所要求时间慢 0.1 秒.

4.24 某汽车经销处分派 20 个职员销售新款汽车. 表 4.33 是三个月内销售员所销售汽车数的分组"大于等于式"累积百分数分布. (a) 如果表中第一列中的数是组下界，请给出第一组(下界为 0 的组)的组限、组中值和组距；(b) 有多少销售员恰好售出 7 辆汽车？(c) 有多少销售员售出的汽车小于等于 7 辆？(d) 售出至少 9 辆汽车的销售员所占的比例为多少？(e) 在这三个月内这 20 个销售员总共售出多少辆汽车？

表 4.33

每个销售员所销售汽车数	累积百分数
大于等于 0	100
大于等于 3	90
大于等于 6	75
大于等于 9	75
大于等于 12	45
大于等于 15	20
大于等于 18	5
大于等于 21	0

解 ☞　(a) 组限是 0—2；组中值是 1；组距是 3

(b) 从表 4.33 可知，有 75%的人售出 6 辆或更多汽车，售出 9 辆或更多汽车的人也是 75%，由于这是一个"大于等于式"累积分布，我们是从大累积到小的；同时由于"销售汽车量大于等于 9"和"销售汽车量大于等于 6"的百分数相等，所以我们可以推出没有销售员恰好售出 6 辆、7 辆和 8 辆汽车.

(c) 由表可知，销售汽车量小于或等于 5 的销售员有(100%－75%＝25%)，即 5 个人；而且，没

有销售员正好售出 6 辆和 7 辆汽车. 所以有 5 个销售员售出的汽车小于等于 7 辆.

(d) "售出至少 9 辆汽车"也即为"销售汽车量大于等于 9",所以从表中可知,售出至少 9 辆汽车的销售员的所占的比例为 75/100 = 0.75.

(e) 因为所给数据是分组数据,所以不能从表 4.33 中得出这一问题的准确解. 例如,20 个人中的 5%,即 1 人,售出 18—20 辆汽车,我们并不能由此得到此人准确的销售量是 18,19 或 20. 尽管从表 4.33 中不能得出这一问题的准确解,但是,我们可以合理地假定某组中所有值均集中在此组中值位置. 在确定了每组的频数后,我们可以得到销售总量的一个估计值为:

估计销售总量 = (1×2)+(4×3)+(7×0)+(10×6)+(13×5)+(16×3)+(19×1) = 206

注:在第六章和第七章中将会看到,当由于只有分组数据而无法得到精确计算值时,常用这种假定某组中所有值均集中在组中值位置的方法来近似其他的描述性测量.

补充习题

阵列和极差

4.25 在一个花样滑冰比赛中,8 个评委给某个男溜冰者的艺术印象得分为:6.0,5.5,5.6,5.9,6.0,5.9,5.7,5.7. 将这些得分表示为升序阵列,并计算其极差.

答案 这些艺术印象得分是评价表演的美感的主观尺度,取值从 0.0(很不好)到 6.0(很好),是次序测量尺度. 所以,无法计算其极差(见习题 4.2);但可以表示为升序阵列:5.5,5.6,5.7,5.7,5.9,5.9,6.0,6.0.

4.26 将附表 A.2 中的女生头发颜色(第 7 列)排成降序阵列,并计算极差.

答案 头发颜色是名义测量尺度(见 2.4 节),所以既不能对其进行从大到小排序(即排成降序阵列),也无法计算它们的极差.

4.27 图 4-1 中的 Colorado 州的地图中记录了六月的某一天 12 个城市的最高温度(以℉为单位). 如可能,请将这些测量值排成升序阵列,并计算极差.

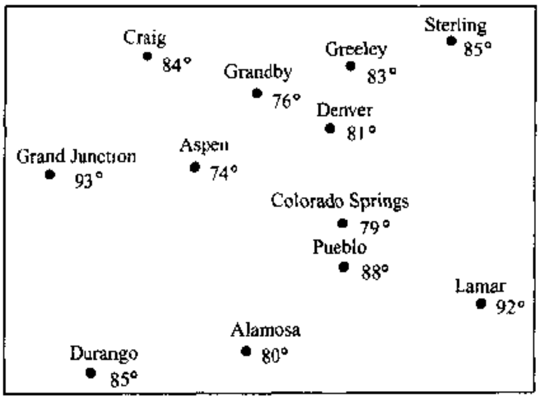


图 4-1

答案 以华氏为单位的温度测量值是间隔测量尺度(见 2.6 节),所以阵列和极差都是可以得到的. 其升序阵列为:74,76,79,80,81,83,84,85,85,88,92,93;极差为:93℉−74℉=19℉.

4.28 一次男子滑翔比赛中的前十名选手的成绩以比赛次序为序记录如下(分:秒.厘秒):1:24.77,1:26.23,1:26.03,1:26.82,1:25.61,1:24.91,1:23.38,1:25.80,1:26.65,1:24.07. 如可能,请先将这些比赛成绩转换成以秒为单位的值,并近似到秒,然后排成升序阵列,并计算极差.

答案 这些比赛成绩是连续比例测量尺度(见 2.8 节),所以阵列和极差都是可以得到的. 转换成秒为单位后的升序阵列为:83,84,85,85,86,86,86,86,87,87;极差为:87 秒—83 秒=4 秒.

分布:频数、相对频数和百分数

4.29 对于表 4.3,计算(a) $\sum_{i=1}^{k}(f_i/N)$,(b) $\sum_{i=1}^{k}[(f_i/N)\times 100]\%$

答案 (a) 1.0,(b) 100

4.30 经济学家为了研究某城市汽油价格的变化趋势,从该城市的 40 个加油站中得到简单随机抽样,以确定每个加油站中每加仑无铅汽油的价格(以美元为单位,是一个离散比例测量尺度). 以下是以升序阵列表示的抽样结果:1.05,1.07,1.07,1.07,1.08,1.10,1.10,1.10,1.10,1.10,1.10,1.10,1.10,1.10,1.10,1.10,1.10,1.11,1.11,1.11,1.11,1.11,1.11,1.11,1.11,1.11,1.11,1.12,1.12,1.12,1.12,1.12,1.13,1.13,1.13,1.13,1.14,1.14,1.14,1.15. 将这一阵列作成汇总表,表中各列为价格、计数、频数、相对频数和百分数.

答案　表 4.34 是完成后的汇总表.

表 4.34

价格($) x_i	计数	频数 f_i	相对频数 f_i/n	百分数 $[(f_i/n)\times100]\%$
1.05	|	1	0.025	2.5
1.06		0	0.000	0.0
1.07	|	3	0.075	7.5
1.08	|	1	0.025	2.5
1.09		0	0.000	0.0
1.10	卌 卌 ||	12	0.300	30.0
1.11	卌 卌	10	0.250	25.0
1.12	卌	5	0.125	12.5
1.13	||||	4	0.100	10.0
1.14	|||	3	0.075	7.5
1.15	|	1	0.025	2.5
$\sum$		40	1.000	100.0%

4.31 某工程师要研究在 60 层的大厦中一种新型空调系统的性能. 为了评价大厦内的温度,空调系统设置为 72℉时,抽取了 20 个房间为简单随机样本,温度精确到℉. 以下是以升序阵列表示的抽样结果:69,69,70,70,70,70,72,72,72,72,72,72,72,72,72,72,73,73,74,74. 将这一阵列作成汇总表,表中各列为温度、计数、频数、相对频数和百分数.

答案　表 4.35 是完成后的汇总表.

表 4.35

温度(℉) x_i	计数	频数 f_i	相对频数 f_i/n	百分数 $[(f_i/n)\times100]\%$
69	||	2	0.1	10
70	||||	4	0.2	20
71		0	0.0	0
72	卌 卌	10	0.5	50
73	||	2	0.1	10
74	||	2	0.1	10
$\sum$		20	1.0	100%

分组分布

4.32 在某杂志的文章中,有一分组频数分布,如表 4.36 所示. 第一列(重量,以 kg 为单位)是每一组的组中值. 第二列(频数)是每组的频数. 由给出的信息,在表中添加组限、组界、组距和百分数各列.

答案　表 4.37 为完成后的分布表.

表 4.36

重量(kg) m_i	频数 f_i
0.3	3
0.8	13
1.3	42
1.8	15
2.3	7
$\sum$	80

表 4.37

重量(kg) m_i	组限	组界	组距	频数 f_i	百分数 $[(f_i/n)\times 100]\%$
0.3	0.1—0.5	0.05—0.55	0.5	3	3.75
0.8	0.6—1.0	0.55—1.05	0.5	13	16.25
1.3	1.1—1.5	1.05—1.55	0.5	42	52.50
1.8	1.6—2.0	1.55—2.05	0.5	15	18.75
2.3	2.1—2.5	2.05—2.55	0.5	7	8.75
$\sum$				80	100.00%

4.33 表 4.22 中为统计学班级学生的期末考试成绩. 在考试前,已给出考试成绩的相应的字母等级:(90—99)为 A,(80—89)为 B,(70—79)为 C,(60—69)为 D,(60 以下)为 F. 请问这个班级中得到每一等级的百分数为多少?

答案 成绩为 A 的学生的百分数为(26.5625+7.8125=34.3750)%,

成绩为 B 的学生的百分数为(17.1875+14.0625=31.2500)%,

成绩为 C 的学生的百分数为(4.6875+9.3750=14.0625)%,

成绩为 D 的学生的百分数为(4.6875+7.8125=12.5000)%,

成绩为 F 的学生的百分数为(1.5625+6.2500=7.8125)%.

4.34 应用 4.6 节中给出的原则,将附表 A.2 中 64 个学生的身高(第 4 列)整理成分组分布,组距为 1.00,第一组的组中值为 60.00. 分布表中应包括分组区间、组界、组中值以及男生和女生分别的计数和频数各列.

答案 表 4.38 为完成后的分布表.

表 4.38

身高(in)	组界	组中值 m_i(in)	女生		男生	
			计数	频数 f_i	计数	频数 f_i
59.50—60.49	59.495—60.495	60.00	|	1		0
60.50—61.49	60.495—61.495	61.00		0		0
61.50—62.49	61.495—62.495	62.00	|	1		0
62.50—63.49	62.495—63.495	63.00	|	1		0
63.50—64.49	63.495—64.495	64.00	|||	3	|	1
64.50—65.49	64.495—65.495	65.00	卌	5	||||	4
65.50—66.49	65.495—66.495	66.00	||	2	|||	3
66.50—67.49	66.495—67.495	67.00	|	1	||||	4
67.50—68.49	67.495—68.495	68.00	|	1	卌 |	6
68.50—69.49	68.495—69.495	69.00	|	1	卌 卌 ||	12
69.50—70.49	69.495—70.495	70.00		0	卌 ||	7
70.50—71.49	70.495—71.495	71.00		0	||||	4
71.50—72.49	71.495—72.495	72.00		0	|||	3
72.50—73.49	72.495—73.495	73.00		0	|||	3
73.50—74.49	73.495—74.495	74.00		0	|	1
$\sum$				16		48

开端点组分布和不等组距

4.35 表 4.39 中的开端点组频数分布摘自《美国统计摘要》(1995 年)的第 671 页. 表中给出了 1992 年美国不同大小(以英亩为单位)的农场的频数分布. 每种大小的农场的数量(以 1992 年为标志的列)是以 1,000 为单位的. 利用给出的信息,添加以下各列:组限、组界、组距和组中值.

表 4.39

农场大小	1992 (农场数(千))	农场大小	1992 (农场数(千))	农场大小	1992 (农场数(千))
10 英亩以下	166	100—179 英亩	301	500—999 英亩	186
10—49 英亩	388	180—259 英亩	172	1,000—1,999 英亩	102
50—99 英亩	283	260—499 英亩	255	2,000 英亩及以上	71

答案　表 4.40 为添加各所需列后的频数分布表.

表 4.40

农场大小	组限	组界	组距	组中值 m_i(英亩)	1992 (农场数(千)) f_i
10 英亩以下	?—9	?—9.5	?	?	166
10—49 英亩	10—49	9.5—49.5	40	29.5	388
50—99 英亩	50—99	49.5—99.5	50	74.5	283
100—179 英亩	100—179	99.5—179.5	80	139.5	301
180—259 英亩	180—259	179.5—259.5	80	219.5	172
260—499 英亩	260—499	259.5—499.5	240	379.5	255
500—999 英亩	500—999	499.5—999.5	500	749.5	186
1,000—1,999 英亩	1,000—1,999	999.5—1,999.5	1,000	1,499.5	102
2,000 英亩及以上	2,000—?	1,999.5—?	?	?	71
$\sum$					1,924

4.36　将附表 A.2 中 64 个家庭的收入(第 6 列)分成与表 4.24 中相同的 9 个组. 分布表中应包括分组区间、计数、频数和百分数.

答案　表 4.41 为完成后的频数分布表.

表 4.41

家庭收入	计数	频数 f_i	百分数 $[(f_i/n)\times 100]\%$
$5,000 以下		0	0.000
$5,000—$9,999		0	0.000
$10,000—$14,999	‖	2	3.125
$15,000—$19,999	‖‖	4	6.250
$20,000—$24,999	卌 卌	10	15.625
$25,000—$34,999	卌 卌 卌 卌 ‖‖	24	37.500
$35,000—$49,999	卌 卌 ‖‖	14	21.875
$50,000—$74,999	卌 |	6	9.375
$75,000 及以上	‖‖	4	6.250
$\sum$		64	100.000%

累积分布

4.37　将表 4.35 中的温度(℉)的频数分布转换成"小于式"累积频数和累积百分数分布.

答案　表 4.42 即为所要求的累积分布.

4.38　将表 4.14 中的产仔数量的频数分布和相对频数分布转换成"大于等于式"累积频数和累积百分数分布.

答案　表 4.43 即为完成后的分布表.

4.39　将表 4.23 中的学生体重的分组频数分布转换成"小于式"累积频数分布.

答案　表 4.44 即为所要求的累积分布. 这一分布是累积到组的上界.

表 4.42

温度(°F)	累积频数	累积百分数
小于 68.5	0	0
小于 69.5	2	10
小于 70.5	6	30
小于 71.5	6	30
小于 72.5	16	80
小于 73.5	18	90
小于 74.5	20	100

表 4.43

产仔数量	累积频数	累积百分数
大于等于 15	50	100
大于等于 16	19	98
大于等于 17	47	94
大于等于 18	42	81
大于等于 19	22	44
大于等于 20	10	20
大于等于 21	4	8
大于等于 22	0	0

4.40　将表 4.23 中学生体重的分组频数分布转换成分组“大于等于式”累积相对频数分布.

答案　表 4.45 即为所要求的累积分布.

表 4.44

体重(lb)	累积频数		
	女生	男生	总数
小于 104.5	0		0
小于 114.5	3	0	3
小于 124.5	11	3	11
小于 134.5	15	1	19
小于 144.5	16	10	26
小于 154.5		21	37
小于 164.5		32	48
小于 174.5		40	56
小于 184.5		44	60
小于 194.5		48	64

表 4.45

体重(lb)	累积相对频数		
	女生	男生	总数
大于等于 104.5	1.00		1.00
大于等于 114.5	0.81	1.00	0.95
大于等于 124.5	0.31	0.94	0.78
大于等于 134.5	0.06	0.92	0.70
大于等于 144.5	0.00	0.79	0.59
大于等于 154.5		0.56	0.42
大于等于 164.5		0.33	0.25
大于等于 174.5		0.17	0.12
大于等于 184.5		0.08	0.06
大于等于 194.5		0.00	0.00

4.41　将表 4.40 中农场大小的开端点组频数分布转换成分组“大于等于式”累积频数和百分数分布,要求在累积中使用组下限.

答案　表 4.46 即为该问题的一种可能解. 应注意的是,因为此分布两端均为开端点组,所以无法确定 0%或 100%的界. 因此,开端点组的频数和百分数应用某种方法单独确定. 由于这些复杂性,很多统计书中建议对于开端点分布不使用累积分布.

表 4.46

农场大小(英亩)	累积频数	累积百分数	农场大小(英亩)	累积频数	累积百分数
(10 以下)	(166)	10 以下占 8.6	大于等于 260	614	31.9
大于等于 10	1,758	91.4	大于等于 500	359	18.7
大于等于 50	1,370	71.2	大于等于 1,000	173	9.0
大于等于 100	1,087	56.5	(2,000 及以上)	(71)	(3.7 2,000 及以上)
大于等于 180	786	40.9			

第五章　描述性统计：统计资料的图形化

5.1　柱状图、线型图和饼状图

统计图是用来反映变量间关系的图形. 这一章主要讨论反映测量变量 X 与频数、相对频数或百分数之间相互关系的图形. 常用的图形有三种：**柱状图**、**线型图**和**饼状图**.

柱状图是在直角坐标系(见 1.20 节)中构造的图形，通常用 X 轴表示自变量，Y 轴表示因变量，而用矩形(或条形)表示变量间的关系. 图 5－1(a) 就是一个频数分布的柱状图，反映了样本中测量值发生的频数(Y 轴)如何随测量变量(X 轴)变化，图中每一个测量值的频数与该值上方的矩形的高成比例. 按照惯例，对于表示经验分布的图形，其坐标轴不用 X、Y 表示.

线型图也是在直角坐标系中构造的图形，它通过一系列用线连接起来的点来表示变量间的关系. 图 5－1(a) 中频数分布的线型图如图 5－1(b)所示，每一个测量值的频数用该值上方点的纵坐标表示，点与点之间用直线连接起来.

柱状图和线型图也可以反映相对频数或百分数的分布，这时须将坐标系中 Y 轴转换成相对频数或百分数.

与柱状图和线型图不同，饼状图不是在直角坐标系中构造的图形. 它是通过一个圆(或饼)中的扇形面积的大小来反映变量间的关系. 本章饼状图用来表示相对频数和百分数分布. 图 5－1(a) 和图 5－1(b) 中所示的频数分布转换成百分数分布后，其饼状图如图 5－1(c) 所示，每一个测量值的百分数与相应的扇形面积成比例.

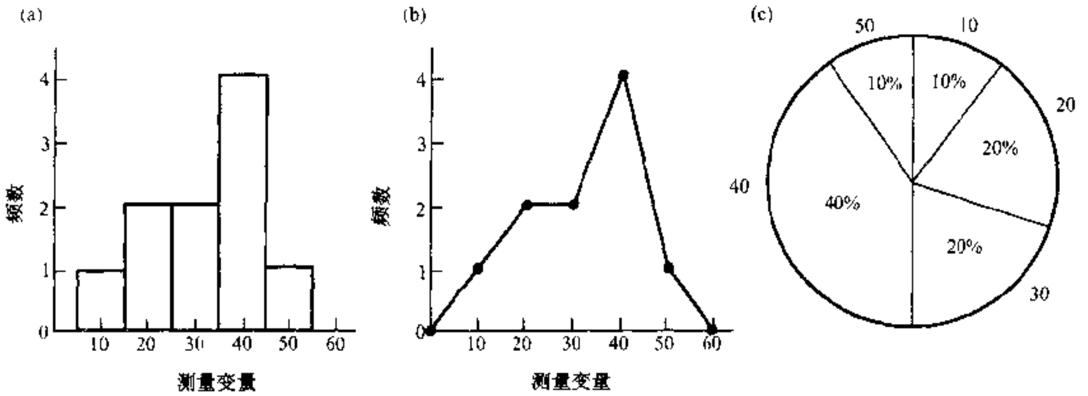


图 5－1

5.2　条形图

条形图是一种专门用于名义水平、次序水平和离散比例水平数据(见 2.4 节，2.5 节和 2.8 节)的柱状图. 它是用矩形的高而不是面积来表示频数、相对频数或百分数，并且各矩形之间留有间隔以表示测量尺度的不明确性或不连续性.

例 5.1　对表 4.18 中男子头发颜色的频数分布，构造一个频数条形图.

解 ☞　该频数分布的频数条形图如图 5－2 所示. 在这个条形图中，测量变量(头发颜色)的类在 X 轴上均匀标出，各类之间留有相等的间隔. 如果测量尺度是有序的(例如，从小到大，从低到高)，那么在 X 轴上各类的次序应保持不变，一般最小的类最靠近 Y 轴. 在这个例子中，变量是名义型的，是无序的，因此可以以任意的次序排列在 X 轴上. Y 轴上的频数尺度以相等的间隔 5 标出. 在每个类的上方是一些宽度相等、高度等于频数尺度上的一个单位的矩形，每一个小矩形表示样本中某个类的一个测量值. 所以每一个类的频数就等于该类垂直方向上小矩形的个数. 一般的条形图

中并不将所有的小矩形表示出来而只是给出最后的大矩形，图 5 - 2 中的小矩形只是用来举例说明条形图的构造过程，类的频数可由条形图顶端在纵轴上的投影读出.

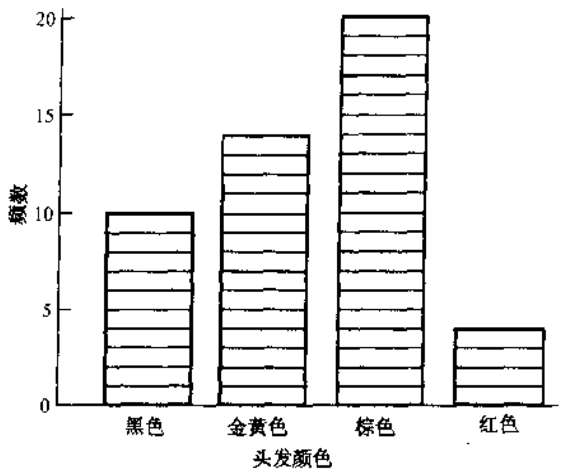


图 5 - 2

矩形之间的间隔有重要的意义. 这些间隔表明测量变量具有统一的、标准的参考单位，但是不连续. 相反的，矩形若是相连的，这种柱状图称为**直方图**(见 5.3 节)，则通常表明测量变量是连续的间隔水平或比例水平测量.

在 Y 轴上构造频数尺度没有固定的规则，但是一般都遵循下列准则：

(1) Y 轴长度大约为 X 轴长度的 60%—75%.

(2) 频数尺度应从 X 轴的零点开始，并以略大于最大频数结束.

(3) 频数尺度上可以有任意多个数，进行任意的划分，但是一般用相等的间隔且间隔为 2—20 之间.

5.3　直方图：未分组数据

正如例 5.1 中所指出的，**直方图**是一种用于连续间隔型和连续比例型数据的柱状图(离散比例水平数据的直方图在习题 5.9 中讨论). 与条形图不同的是，直方图用矩形的面积而不是高度来表示频数、相对频数或百分数，并且各矩形是相连的，形成一个连续结构. 例 5.1 中构造条形图的建议和原则也适用于构造直方图.

例 5.2　将表 5.1 中的数据构造一个**频数直方图**.

表 5.1

长度(cm) x_i	频数 f_i
1.2	2
1.3	7
1.4	10
1.5	12
1.6	10
1.7	7
1.8	2
$\sum$	50

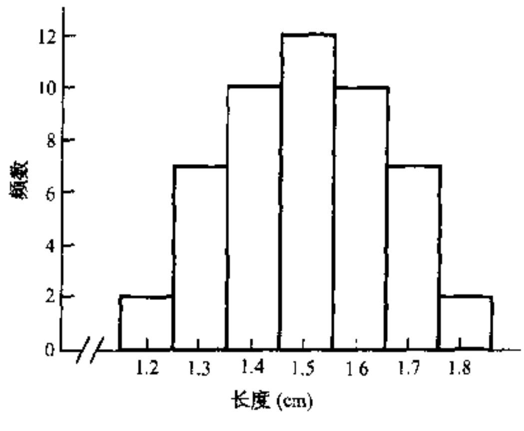


图 5 - 3

解 ☞　该数据的频数直方图如图 5 - 3 所示. 对于连续比例水平数据的未分组频数分布，直方图中的每一个矩形表示一个测量类，矩形的底宽是从对应测量类的下界到上界. 每一个矩形下方的中点数值就是该类的测量值. 这样构造直方图的前提假设是：在每个类内的所有样本测量值在其

蕴含范围是均匀分布的. 如果所有矩形的宽度都相等,那么每一个类的频数与相应矩形的高度和面积都是成比例的.

从这个直方图的中点(X轴上的1.5)作一垂直线,可以看出该垂直线将直方图分成了两个完全相等的对映部分,我们称这样的直方图是**对称**的. 同样可以看出该直方图只有一个峰,所以称为**单峰**的(含有两个峰的直方图称为**双峰**的,含有三个峰的直方图称为**三峰**的,依此类推;将在第六章讨论众数和多峰).

5.4　直方图:分组数据

根据定义(见习题4.14)分组数据都是间隔型或比例型,所以直方图一般用于连续测量变量,而不用于离散比例型数据(见习题5.9). 但是对于分组离散比例型数据,直方图仍然适用.

在分组数据的频数直方图中,每一个矩形表示一个组的频数. 矩形的底边从组下界到组上界,宽度代表该类的组距. 如果每组的底宽相等,则组频数与矩形的面积和高度均成比例. 如此构造直方图的前提假设是:每组中的测量值在其组界内是均匀分布的.

例5.3　用表4.36中的重量的分组频数分布,构造一个频数直方图. 其中,X轴表示组中值(m_i).

解 ☞　该分布的频数直方图如图5-4所示. 图中矩形在X轴上以组中值为中点,但是对于这样的分组数据同样也可以用组界、组限等作为X轴.

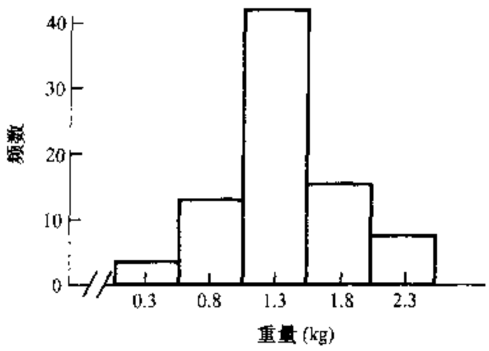


图5-4

5.5　折线图:未分组数据

折线图是用折线表示频数、相对频数或百分数分布的线型图(见5.1节). 对未分组连续数据构造折线图时,我们假定每一类内的所有测量值均落在该类区间的中点值上. 例如,对于**频数折线图**,用来表示每一类的频数的点均在该类的中点处,其高度由垂直的频数尺度表示,然后将这些点用直线相连. 习惯上,我们常沿X轴延长折线两端,并把折线的左端与最小类的前一个类的中点相连,把折线的右端与最大类的后一个类的中点相连. 这样连同X轴就产生了一个由折线构成的封闭的平面图,即折线图.

例5.4　将表5.1中的数据构造一个频数折线图. 并在图中显示该数据的频数直方图(图5-3)与频数折线图的联系.

解 ☞　该数据的频数折线图和频数直方图如图5-5所示. 图中折线是直方图中相邻矩形的顶端中点的连线. 尽管频数折线图和频数直方图联系紧密,但是很少将两个图形放在一起. 当单独构

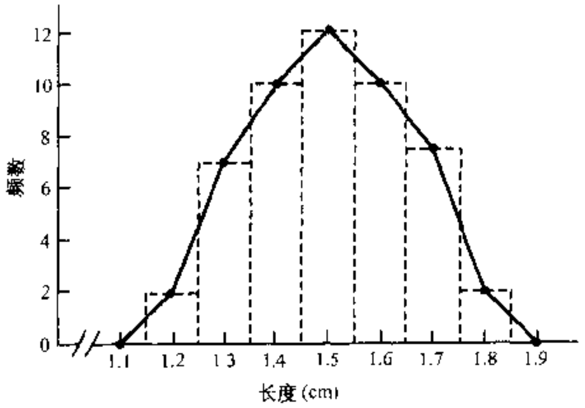


图5-5

造折线图时，先在坐标系中描出各点(x=观测值，y=频数)，然后再将这些点用直线相连．那些描述直方图的特征同样适用于折线图，所以这个折线图是单峰的，对称的.

5.6　折线图:分组数据

对分组数据构造折线图时，假定每组内的所有测量值均落在其组中值上．若组距相等，则代表频数、相对频数、百分数的点都在组中值的正上方，纵坐标为适当的高度．用直线连接这些点，和例 5.4 一样．折线是直方图中相邻矩形的顶端中点的连线．与未分组数据的折线图类似，假定直方图左、右端的分布有相同的组距，将折线分别沿 X 轴的两端延长，构造一个封闭的折线图.

例 5.5　利用表 4.21 中的分组数据构造一个频数折线图和直方图．图中 X 轴上标出组中值 m_i.

解 ☞　该数据的频数折线图和频数直方图如图 5-6 所示.

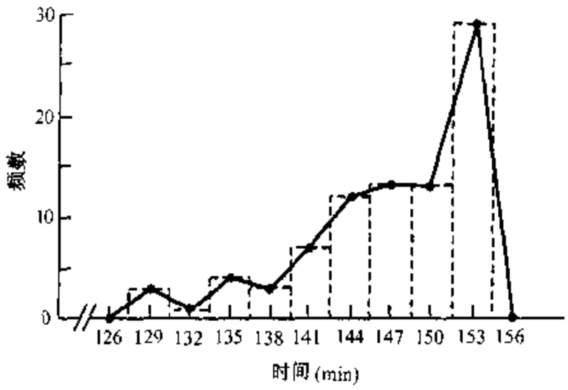


图 5-6

5.7　频数曲线、相对频数曲线和百分数曲线

分组数据的频数折线图通常用来表示组数为 5 至 20 的样本的频数分布．但是，如果样本容量逐步增加，分布的组数越来越多，组距逐步减小，那么折线图会有什么变化？在这种情况下，当样本容量 n 接近总体容量 N 时，频数折线图会变成一平滑曲线，称为**频数曲线**(或**平滑曲线频数折线图**)．类似地，相对频数折线变成**相对频数曲线**，百分数折线变成**百分数曲线**．这种总体的平滑曲线可以用理论的数学曲线近似(或**拟合**)．在后面的章节中可以看到，这些理论曲线是推断统计学中概率决策的基本工具.

5.8　象形图

象形图是一种用具有特殊风格的、简单易认的图象代替矩形来表明统计资料的柱状图.

例 5.6　表 5.2 中的开端点组频数分布摘自**《美国统计摘要》**(1995)的第 471 页．这是 1993 年美国的九个不同收入组中单身男子和单身女子的数量．我们想用**相邻部分相对频数象形图**表示这些数据．为了简化这一表示，先将九个收入组压缩到三个收入组：低收入(低于＄25,000)，中等收入(＄25,000 至＄49,999)和高收入(等于或高于＄49,999)．将表 5.2 的频数分布转换成分成三组的相对频数分布，并作出相对频数象形图.

表 5.2

家庭收入	单身家庭数(以 1000 为单位)	
	男子单身	女子单身
	f_i	f_i
$5,000 以下	704	1,386
$5,000—$9,999	1,257	4,013
$10,000—$14,999	1,324	2,439
$15,000—$19,999	1,106	1,678
$20,000—$24,999	1,080	1,279
$25,000—$34,999	1,639	1,578
$35,000—$49,999	1,217	1,083
$50,000—$74,999	698	520
$75,000 及以上	414	196
$\sum$	9,439	14,172

解 ☞　表 5.3 表示的是相对频数分布,相应的相对频数象形图如图 5－7 所示. 图中,每一个完整的小人,男或女,代表 0.1 个单位的相对频数.

表 5.3

家庭收入	单身家庭数(以 1000 为单位)			
	男子单身		女子单身	
	频数	相对频数	频数	相对频数
	f_i	f_i/N	f_i	f_i/N
低收入	5,471	0.5796	10,795	0.7617
中等收入	2,856	0.3026	2,661	0.1878
高收入	1,112	0.1178	716	0.0505
$\sum$	9,439	1.0000	14,172	1.0000

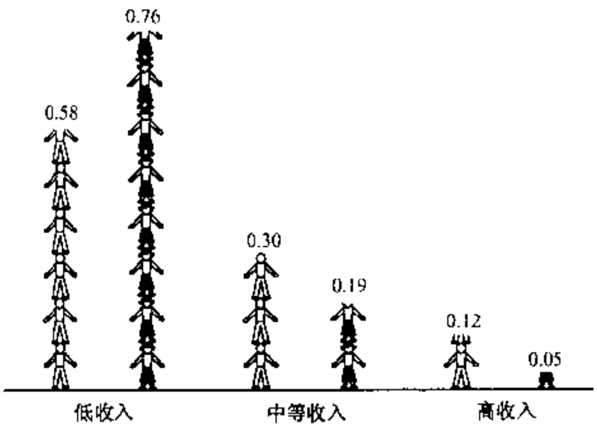


图 5－7

5.9　饼状图

如 5.1 节所述,**饼状图**(或称**圆形图**)是把全圆分成若干扇形部分,以各扇形面积的大小表

示相对频数或百分数的值. 如果分布的类(或组)是有序的,那么一般从 12 点钟位置开始顺时针次序排列各类.

例 5.7　利用表 5.3 的数据分别作出男子和女子的相对频数饼状图.

解 ☞　相对频数饼状图如图 5-8 所示.

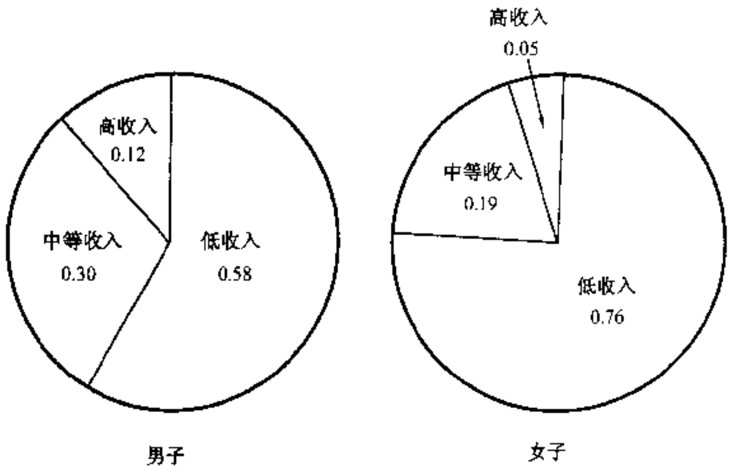


图 5-8

5.10　茎叶表示法

茎叶表示法能快速地、非正式地观察数据集特征. 通过这种方法能初步观察分布的对称性和形状特征,并能帮助确定分组数据分布的最合适的组距. 在通常的茎叶表示法中,有一列垂直的数字,称为**开始部分**,每一个开始部分的旁边有一行水平的数字,称为**叶**. 每一个开始部分和其旁边的叶称为一条**茎**. 数据集中的每一个数在茎叶表示法中都由一个开始部分和一个叶子组成.

茎距是指某茎上能记录的最小值与下一个茎上能记录的最小值之间的距离,它决定数据集中哪些数记录在指定的茎上.

简单的茎叶表示法一般有下列特征:(1)每一条茎都有不同的开始部分,(2)每一个开始部分可以用一位以上的数字表示,但是茎上的每一个叶子只能用一位数字表示.

例 5.8　将下列数字用一位数的开始部分及叶子,茎距为 10 的简单茎叶表示法表示:46,35,37,20,43,15,15,26,45,25,29,13,39,44,21,24,16,40,19,45,30,34,17,39,16,40,31,21,14,42,16,43,22,11,24,25,31,27,40,33

1	5536976461	(10)
2	06591412457	(11)
3	579049113	(9)
4	6354050230	(10)
		(40)

图 5-9

解 ☞　图 5-9 即为所要求的简单茎叶表示图. 直线左边的垂直数字(1,2,3,4)是开始部分,直线右边的水平数字就是叶子. 例如,对于开始部分为 2 的茎,其叶子是 06591412457. 在这里,如我们要求用一位数的开始部分和叶子,开始部分表示数据的十位数字,而叶子表示数据的个位数字,所以开始部分为 2 的茎表示的数据是:20,26,25,29,21,24,21,22,24,25,27. 我们要求茎距为 10,所以开始部分为 1 的茎上应包括数 10 到 19,而下一茎则从数 20 开始.

当所有的数据都已经表示出来时,我们将每一条茎中表示的数据的个数在该茎的最右边用括号表示出来,然后进行汇总. 这样可快速检查是否所有数据都已经记录在茎叶表示图中.

5.11 累积分布曲线图

累积曲线是用来表示连续型累积频数、相对频数或百分数分布(见 4.8 节、4.9 节和 4.10 节)的图形. 例如,**"小于式"累积频数曲线**(或**"小于式"累积频数折线图**)是用来表示连续数据的"小于式"分组或未分组累积分布的线型图(见 5.1 节).

对于未分组的数据,"小于式"累积是累积到类的蕴含范围的上界为止(见 4.8 节),所以将这些上界在 X 轴上标出,Y 轴则表示累积频数. 图中各点以类的蕴含上界为横坐标,以直到这个上界为止的累积频数为纵坐标描出,并用直线将各点连接起来. 最后把折线的最左端的点与 X 轴上最小类的前一个类的中点相连,就得到了"小于式"累积频数曲线,它是从左到右逐渐上升的图形.

"大于等于式"频数累积曲线(或**"大于等于式"累积频数折线图**)是在直角坐标系上用来表示"大于等于式"连续数据的分组或未分组累积分布的线型图,它的构造方法与"小于式"累积曲线的构造方法相类似. 对于未分组的数据,"大于等于式"累积是累积到类的蕴含范围的下界为止(见 4.9 节),所以将这些下界在 X 轴上标出,Y 轴同样表示累积频数. 图中各点以类的蕴含下界为横坐标,以直到这个下界为止的累积频数为纵坐标描出,并用直线将各点连接起来. 类似地,把折线的最右端的点与 X 轴上最大类的后一个类的下界相连,就得到了"大于等于式"累积频数曲线,它是从左到右逐渐下降的图形.

例 5.9 将表 4.10 中的"小于式"累积频数分布表示成"小于式"累积频数曲线.

解 ☞ 图 5-10 即为要求的"小于式"累积频数曲线. 从图中看出,1.25 kg 上方的点的高度是 6,说明样本中有 6 个测量值小于 1.25 kg. 从 1.55 kg 上方的点可知,样本中所有 20 个测量值都小于 1.55 kg. 1.05 kg 到 1.15 kg 的"小于式"累积频数没有变,这个事实表明小于 1.05 kg 和小于 1.15 kg 的数都是 2 个,没有测量值是 1.1 kg.

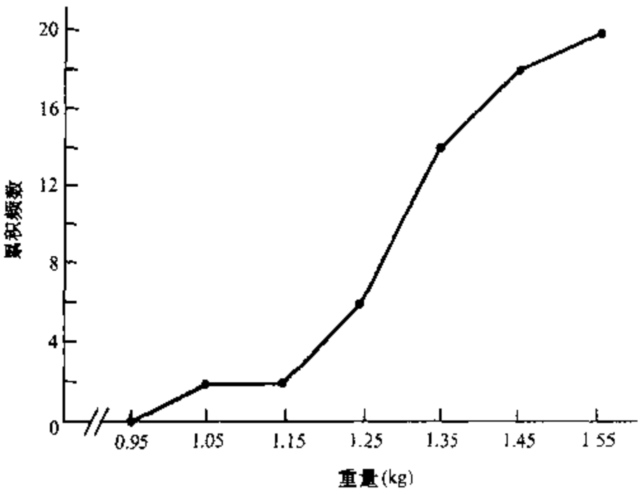


图 5-10

例 5.10 将表 4.29 中的"大于等于式"累积频数分布表示成"大于等于式"累积频数曲线.

解 ☞ 图 5-11 即为要求的"大于等于式"累积频数曲线. 从图中看出,1.25 kg 上方的点的高度是 14,说明样本中有 14 个测量值大于等于 1.25 kg. 1.55 kg 上方的点表明样本中没有测量值大于等于 1.55 kg. 并且 1.05 kg 到 1.15 kg 的"大于等于式"累积频数没有变,这个事实表明样本中没有测量值是 1.1 kg.

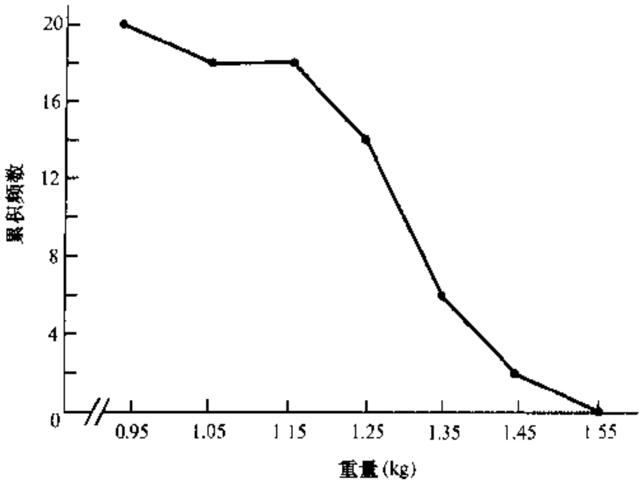


图5-11

习题解答

条形图

5.1 表4.16是女生的学期论文成绩的频数分布,图5-12是这一分布的条形图.请指出该条形图中的错误.

解 ☞ 该条形图中有5个错误(见例5.1):

(1)表4.16中学期论文成绩是从低到高排列的.所以在条形图中从左到右类的排列顺序应该是:F,D,C,B,A.

(2)由于测量尺度是有序的,不是具有统一的标准的参照单位的连续变量,所以图中矩形之间应该留有间隔.

(3)纵轴上尺度的标出只是为了说明每类的频数,这是不正确的,应划分为相等的间隔.

(4)矩形的宽度不相等.图中每个矩形的宽度是其高度的1/3.

(5)通常要求纵轴长度大约为横轴长度60%~75%,但是该图中纵轴和横轴的长度相等.

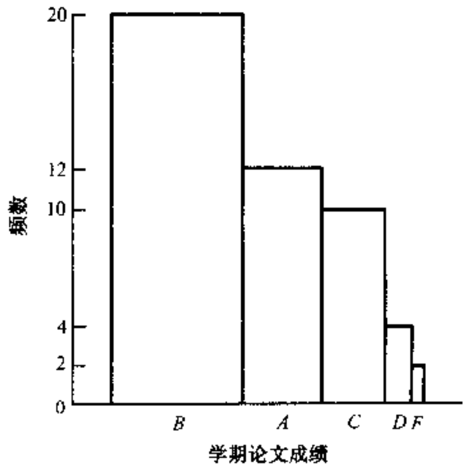


图5-12

5.2 构造上题中女生学期论文成绩的条形图(更正图5-12中的所有错误),并在同一个图中表示频数和相对频数.

解 ☞ 图 5-13 即为所要求的条形图，其中在右侧增加的纵轴表示相对频数，它的刻度是相等的间隔 0.1. 相对频数尺度是由频数尺度的坐标协调产生的：首先将相对频数尺度值转换成等价的频数，然后在频数尺度上找到这些频数的位置，最后，在右侧尺度的相同位置标出相对频数值. 例如，相对频数 0.1 等价于频数 4.8(0.1 × 48)，因此相对频数轴上的刻度 0.1 与频数轴上的刻度 4.8 有相同的高度.

条形图可以在只有左侧纵轴(频数、相对频数或百分数)的直角坐标系构造，也可以在有两个平行纵轴的直角坐标系构造.

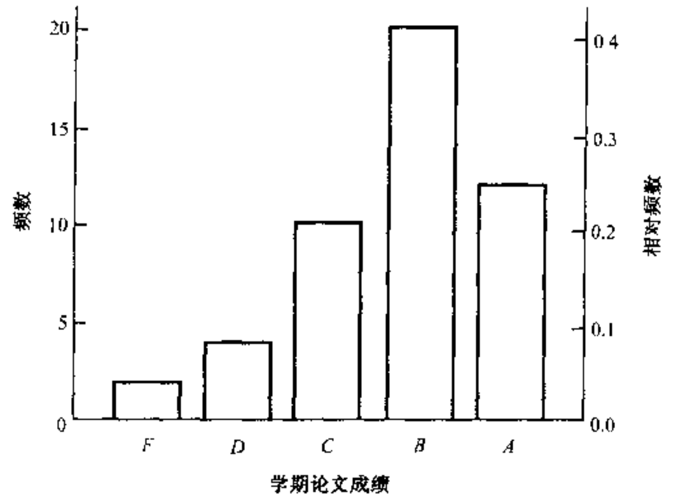


图 5-13

5.3 根据表 4.17 中女生头发颜色的数据，构造一个**水平方向的百分数条形图**.

解 ☞ 图 5-14 即为所要求的条形图. 名义型数据的水平方向的条形图中，类的排列次序通常遵循如下原则：最长的矩形排列在最上端，然后向下逐步递减到最短的矩形；或者最长的矩形排列在最下端，然后向上逐步递增到最短的矩形. 矩形的长度表示的数可以由矩形右边的数字表示或者从图中横轴上的刻度读出. 水平方向的条形图中矩形之间同样留有间隔以表示类边界的不明确性或不连续性.

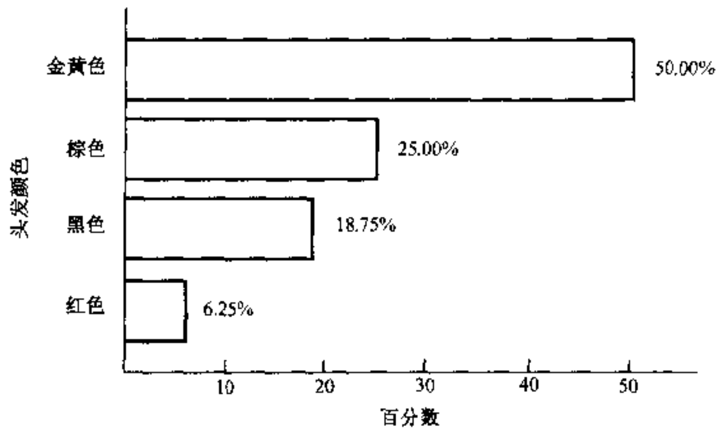


图 5-14

5.4 将表 4.17 中女生头发颜色的数据和表 4.18 中男生头发颜色的数据放在一起，构造一个**成分部分频数条形图**和一个**相邻部分频数条形图**.

解 ☞ 图 5-15(a) 即为所要求的成分部分频数条形图(也称**分量条形图**)，图 5-15(b) 即为所要求的相邻部分频数条形图. 在成分部分频数条形图中，每一个矩形的总长度表示一个头发颜色类的总频数，每个矩形都分成两部分，分别表示男生和女生的频数. 在相邻部分频数条形图中，每一类的男生和女生的频数都用成对的矩形表示.

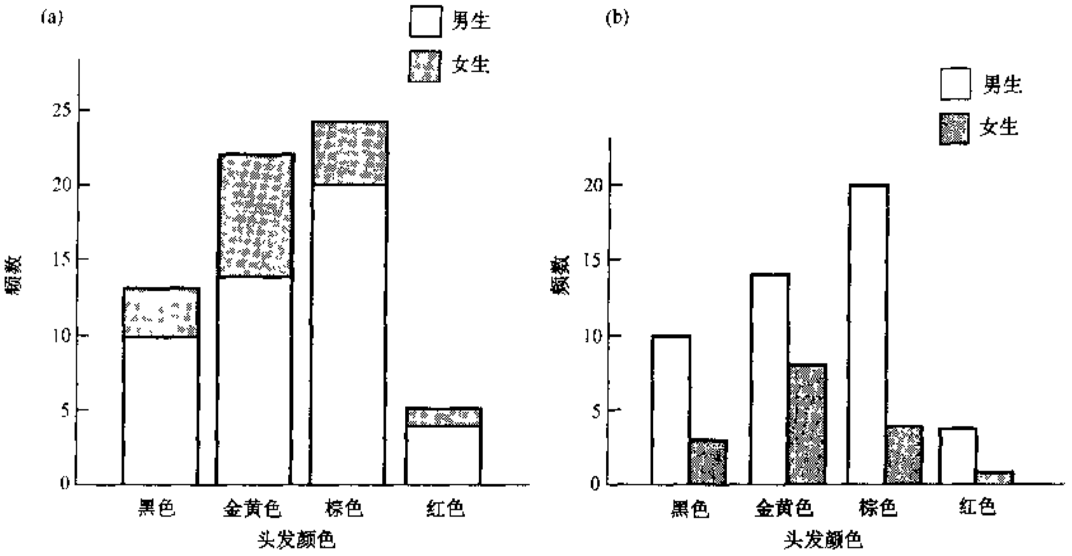


图 5-15

直方图:未分组数据

5.5 根据表 5.4 中的数据,构造一个**相对频数的直方图**.

表 5.4

重量(g) x_i	频数 f_i	相对频数 f_i/n
14	2	0.0222
15	2	0.0222
16	4	0.0444
17	18	0.2000
18	24	0.2667
19	35	0.3889
20	5	0.0556
$\sum$	90	1.0000

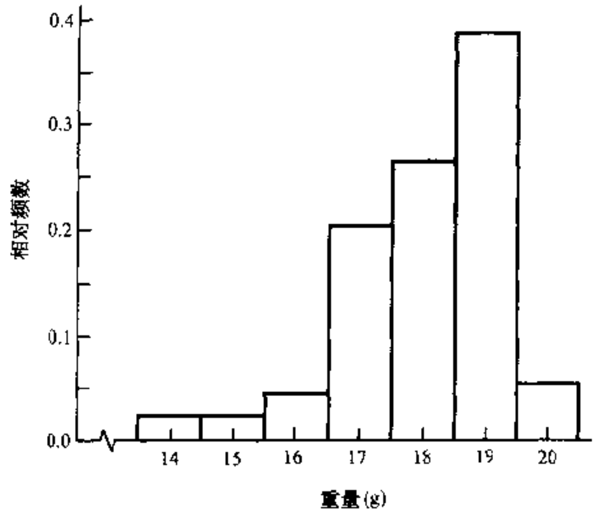


图 5-16

解 ☞ 图 5-16 即为所要求的直方图. 与例 5.2 中相同,对于连续比例型数据使用直方图. 此直方图也是单峰的,但是与图 5-3 不同的是,它不是对称的,它有一个低的相对频数尾部沿着 X 轴的负方向向左边延伸. 由于它不是对称的,所以它也称为**偏斜**的,在这个直方图中,称为**左偏斜或负偏斜**.

5.6 根据表 5.5 中的数据,构造一个同时表示频数和相对频数的直方图.

解 ☞ 图 5-17 即为所要求的直方图. 由于数据是连续间隔型,所以直方图同样适用. 注意到该直方图中有一个缺口,这是用来表示测量值 105°F的缺失. 这个缺口说明尽管直方图适用于连续型测量值,但是直方图本身并不一定是连续结构. 同样可用习题 5.2 中的方法来构造双纵坐标的频数和相对频数尺度. 此直方图是单峰的,但是它有一个尾部,沿着 X 轴的正方向向右边延伸,这种情况称为**右偏斜或正偏斜**.

表 5.5

温度(°F) x_i	频数 f_i	相对频数 f_i/n
100	10	0.10
101	45	0.45
102	25	0.25
103	10	0.10
104	5	0.05
105	0	0.00
106	3	0.03
107	2	0.02
$\sum$	100	1.00

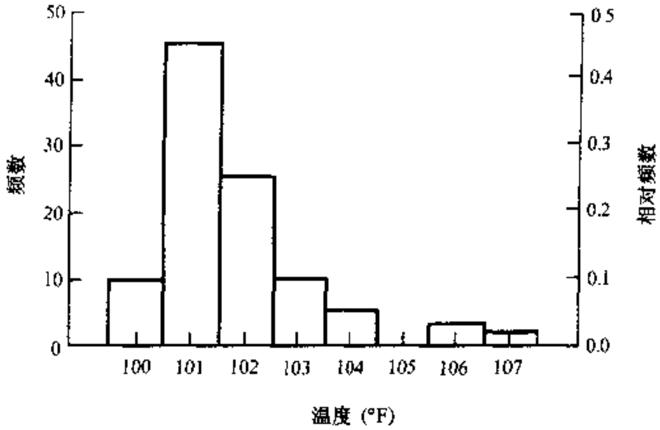


图 5-17

5.7 图 5-3 和图 5-16 中,在靠近 Y 轴的 X 轴上都有一个分割号,用来说明最小测量值与 0 之间有间断. 图 5-3 中使用的分割符号是(-//-),而图 5-16 中使用则是(⌃V). 在图 5-17 中,最低温度与 0 之间同样有间断,但是图中并没有分割符号. 请问哪一种表示间断的方法是正确的?

解 ☞ 类似的,还有其它一些符号,例如(#),都经常用来表示间断. 图 5-17 中并没有采用符号来表示间断,而是在靠近 Y 轴的地方开始作直方图,这也是经常采用的一种方法.

5.8 图 5-3 的直方图中的矩形具有相等的宽度,所以每一个类的频数与相应矩形的高度和面积都是成比例的. 类似地,图 5-16 中矩形的宽度相等,这说明每一个类的相对频数与相应矩形的高度和面积都是成比例的. 请问下列各图中直方图的总面积(即所有矩形的面积之和)与什么成比例:(a) 图 5-3,(b) 图 5-16,(c) 图 5-17?

解 ☞ (a) 总面积与频数之和成比例.

(b) 总面积与相对频数和(即 1.000)成比例.

(c) 对于频数尺度,总面积与频数之和成比例. 对于相对频数尺度,总面积与 1.000 成比例.

5.9 用 40 只白鼠进行试验,让它们学会走一个复杂的迷宫. 如果白鼠在迷宫中没有出现错误(即错误转弯的次数为 0),那么认为它达到了**标准**,学会了走迷宫. 表 5.6 是这次试验的试验次数频数分布. 请问对于这个数据应使用条形图还是直方图?

表 5.6

达到标准的试验次数 x_i	频数 f_i
15	1
16	3
17	9
18	15
19	8
20	2
21	2
$\sum$	40

解 ☞ 例 5.1 中指出,矩形之间的间隔表明测量变量不是具有统一的、标准的参考单位的连续变量. 表 5.6 中的数据是离散比例水平的,因此解释这个原则时进退两难. 这里测量变量(达到标准的试验次数)不是连续的,但是测量尺度处于较高水平(比例水平),因此变量具有统一的、标准的参考单位(数 1). 由于这种进退两难的状况,所以两种柱状图都可以使用. 条形图用来强调变量的离散性质,直方图则强调数据集处于比名义水平和次序水平更高的水平.

还存在另外一个重要原因使得这类数据可以使用直方图,这与参数统计学的假定有关. 3.13 节中提到,参数统计方法是间隔水平和比例水平的首选方法,但是前提是满足参数假定条件. 其中一个假定是**正态分布假定**,包括要求测量变量是连续的. 参数方法不能用于离散比例水平测量值. 然而幸运的是统计学家已经认识到,非连续性不是比例水平数据的严重问题,因此他们建议将离散比例水平数据按连续的处理,以使用参数方法. 这就是对这类数据使用直方图的另一个重要原因. 许多统计书为了解决这个矛盾,声明尽管测量变量是离散的,但是存在一个潜在的连续假设变量. 例如,测量变量(达到标准的试验次数)是离散的,但是潜在的假设变量"学习"是连续的.

5.10　根据表 5.6 中的数据,构造一个杆图.

解 ☞　图 5-18 即为所要求的杆图. 这种表示频数的图形是条形图的一种形式,图中细线的高度与频数成比例,同时假定此细线没有宽度或面积. 这又是表述离散数据的一种方式,它的目的是强调测量变量的不连续性.

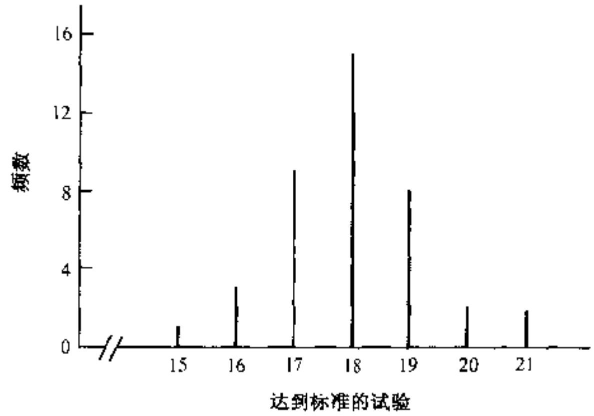


图 5-18

直方图:分组数据

5.11　利用表 4.21 中的分组数据构造一个相对频数直方图,并在 X 轴上标出组界.

解 ☞　图 5-19 即为所要求的直方图. 此直方图的最大频数在右端,并向左边有一个频数逐渐变小的下坡,这种图形是负偏斜的,称为 **J 形的**. 若数据的特征相反(即最大频数在左端,并向右边有一个频数逐渐变小的下坡),这种正偏斜的分布则称为**反 J 形的**.

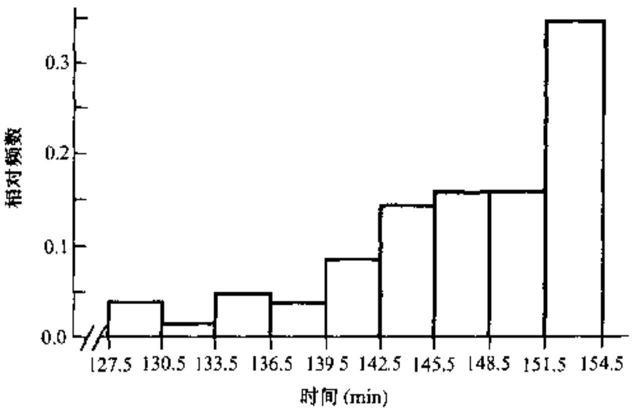


图 5-19

5.12　指出图 5-20 所示的频数直方图的错误,图中 X 轴上已标出组界,组频数与矩形高度成比例.

解 ☞　这是一个具有不等组距的分组数据的例子. 图中左边第一组的组界是 10.5 kg 到 25.5 kg,所以组距为 15;而其他六组的组距均为 5. 对于这样的不等组距,如果矩形高度与频数成比例,那么矩形的面积与频数就不成比例. 而在频数直方图中,矩形的面积应与频数成比例. 这种不成比例性就是此图的错误所在. 例如,图中组界是 10.5 kg 到 25.5 kg 的组与组界是 30.5 kg 到 35.5 kg 的组的频数的比例为 30/30 或 1/1,而它们的面积之比为 3/1. 对于这一问题公认的解决方法如下:

(1)选取一个**标准参考宽度**,通常是分组分布中的最小组距.

(2)用每个组距除以参考宽度,并将结果求倒数.

(3)将每组的频数乘以该组的倒数,得到一个乘积,称为**参考宽度的频数**,又称**频数密度**.

(4)构造直方图，图中纵轴表示频数密度.

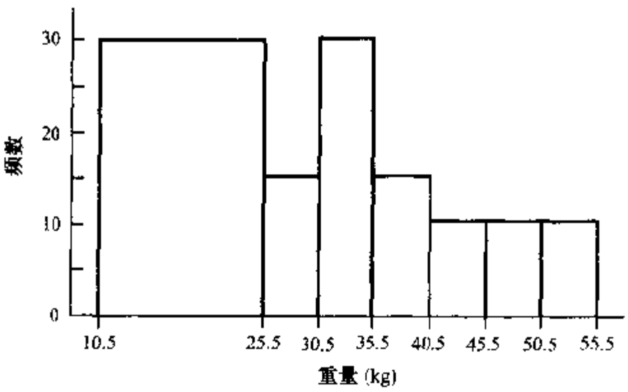


图 5-20

5.13 利用上题给出的方法，更正图 5-20 频数直方图中的错误.

解 (1)图中最小的组距为 5，即为参考宽度.

(2)用每个组距除以 5，则第一组(10.5 kg 到 25.5 kg)为 15/5=3，倒数为 1/3；其他六组为 5/5=1，倒数为 1.

(3)将每组的频数乘以该组的倒数(见表 5.7)，得到参考宽度的频数，这里是每 5 kg 的频数.

(4)图 5-21 即为完成后的直方图. 图中纵轴表示每 5 kg 的频数.

表 5.7

组界	(组频数	×	该组倒数)	=	每 5kg 的频数
10.5—25.5	(30	×	1/3)	=	10
25.5—30.5	(15	×	1)	=	15
30.5—35.5	(30	×	1)	=	30
35.5—40.5	(15	×	1)	=	15
40.5—45.5	(10	×	1)	=	10
45.5—50.5	(10	×	1)	=	10
50.5—55.5	(10	×	1)	=	10

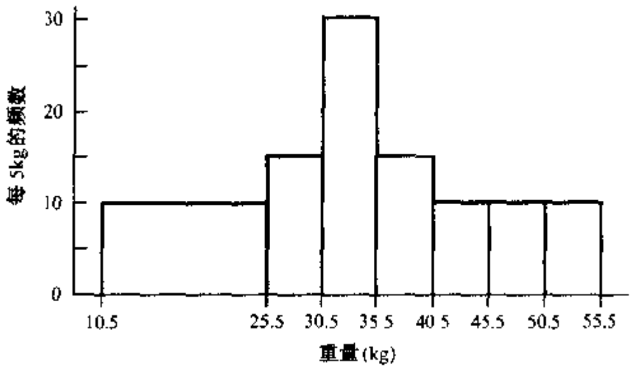


图 5-21

经过这样处理，每个矩形的面积与频数成比例，而高度则与每 5 kg 的频数成比例. 为了得到面积与频数之间的关系，考虑两组(10.5 kg 到 25.5 kg)与(30.5 kg 到 35.5 kg)，它们的频数比例为 30/30 或 1/1. 在图 5-21 中，它们的矩形面积比例为 1/1，而不是图 5-20 中的 3/1. 我们还可以看出，图 5-20 中的直方图是双峰的、正偏斜的；而更正后的图 5-21 中的直方图是单峰的、对称的.

5.14 利用表 4.25 中家庭收入的分组数据构造一个百分数直方图,并在 X 轴上标出所用的测量尺度.

解 ☞ 这个例子中的数据不仅具有不等组距,而且还具有开端口组.为了构造正确的直方图,需要将百分数转换成参考宽度的百分数,但是这一运算仅限于有确定组界的组.由于最小组距为 \$5,000,我们取其为参考宽度,并且计算每 \$5,000 的百分数.计算步骤如表 5.8 所示,图 5-22 即为所要求的直方图.应注意的是有关开端口组的信息已在直方图上写出.

表 5.8

家庭收入	(组百分数	×	该组的倒数)	=	每 \$5,000 的百分数
\$5,000 以下	(4.5383	×	未定义)	=	?
\$5,000—\$9,999	(9.7490	×	1/1)	=	9.7490
\$10,000—\$14,999	(9.2228	×	1/1)	=	9.2228
\$15,000—\$19,999	(8.5668	×	1/1)	=	8.5668
\$20,000—\$24,999	(8.3444	×	1/1)	=	8.3444
\$25,000—\$34,999	(14.7446	×	1/2)	=	7.3723
\$35,000—\$49,999	(16.2614	×	1/3)	=	5.4199
\$50,000—\$74,999	(16.0977	×	1/5)	=	3.2195
\$75,000 及以上	(12.4749	×	未定义)	=	?

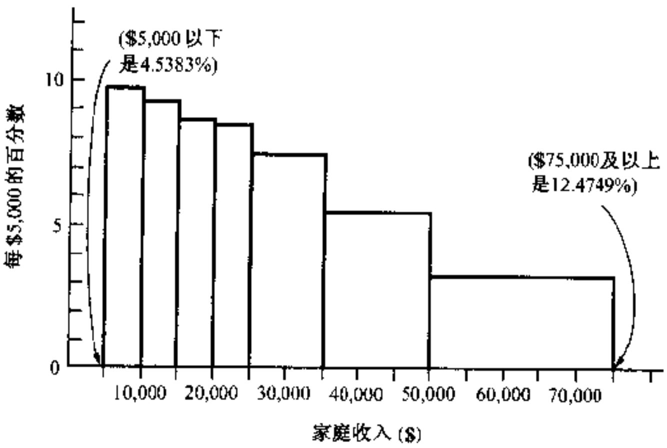


图 5-22

折线图:未分组数据

5.15 利用表 5.4 中的数据构造一个**相对频数折线图**,并在图中做出与之相关的直方图.

解 ☞ 该数据的折线图和直方图如图 5-23 所示.同样假设每类的所有值均集中在该类蕴含范围

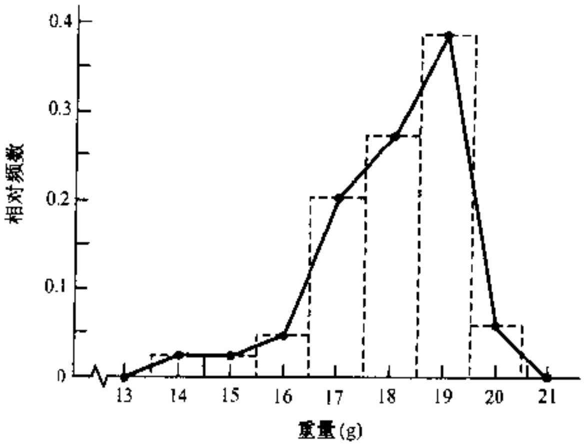


图 5-23

的中点,这里中点上方的那些点的高度表示相对频数,用线段连接这些点,并延伸到 X 轴. 与频数折线图一样,相对频数折线图也可以由其相关直方图来构造,或者直接在坐标系中描点来得到. 这个折线图,与其直方图一样,是单峰的、负偏斜的.

5.16 由例 5.2 和习题 5.8 可知,图 5-5 中的直方图的总面积与 $n=50$ 成比例,图 5-23中的直方图的总面积与 1,000 成比例. 这些图中的折线图的总面积与什么成比例?

解 ☞ 图 5-5 中的直方图和折线图的面积相等,因此折线图围成的面积与 $n=50$ 成比例. 图 5-23中折线图和直方图也具有相等的面积,因此折线图围成的面积与 1,000 成比例. 事实上,如果在相同的坐标系中对相同的分布构造直方图和折线图,那么通常它们的面积都相等. 为了说明这点,将图 5-5 中的前三个类放大,见图 5-24.

从图 5-24 中可以看出,沿着直方图矩形的左侧边,折线左边的三角形与直方图左边的三角形是全等三角形. 考虑第一对三角形(浅色阴影),有一个三角形在折线下方,但是在直方图外面,而另一个三角形在直方图下方,但是在折线上面. 这种关系对于第二对三角形(深色阴影)也是成立的. 同样对于图 5-5 中所有其它对三角形也是成立的. 因此,每一个直方图上的三角形面积,是折线图所没有的,而在折线图中添加了一个相等的三角形面积. 因此,直方图和折线图的总面积总是相等的. 从这也可看出,将折线沿 X 轴扩展形成封闭的图形,也是为了得到相等的面积.

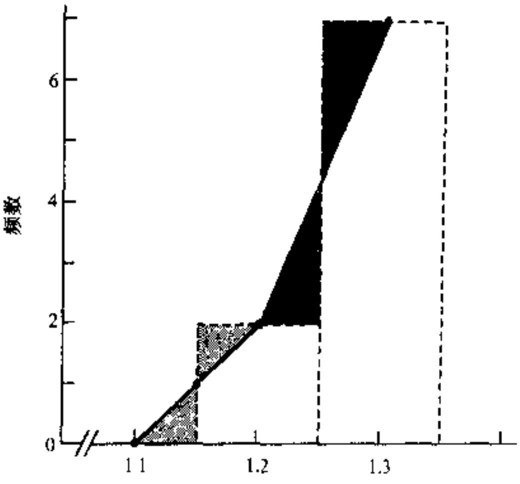


图 5-24

5.17 名义水平、次序水平和离散比例水平的数据可以构造折线图吗?

解 ☞ 名义水平(图 5-2)和次序水平(图 5-13)的数据是不连续的,没有统一的、标准的参照单位,因此需使用条形图(矩形间具有间隔以说明测量变量的不连续性),所以这类数据不能使用连续的线型图表示. 然而却允许使用折线图表示离散比例数据,其原因与习题 5.9 中的讨论相同.

折线图:分组数据

5.18 对表 4.22 中的分组离散比例数据构造一折线图,同时表示频数和百分数、不给出有关的直方图,而直接在坐标系中描点,并在 X 轴上标出组中值(m_i).

解 ☞ 所要求的折线图见图 5-25.

5.19 根据习题 5.12 的信息,将每 5kg 频数的直方图(图 5-21)转换成每 5kg 频数的折线图. 并给出相关的直方图和折线图,在 X 轴上标出组中值.

解 ☞ 所要求图形见图 5-26. 构造这个折线图的问题是使折线图的面积与图 5-21 中直方图的面积相等. 为了解决这个问题,图 5-26 中的第一个矩形的宽是其它矩形的三倍,将它看成三个宽为 5kg 的矩形的组合,然后用等组距分布构造折线图的相同方法来构造这个折线图.

5.20 图 5-22 中每 \$5,000 百分数的直方图能转换成每 \$5,000 百分数折线图吗?

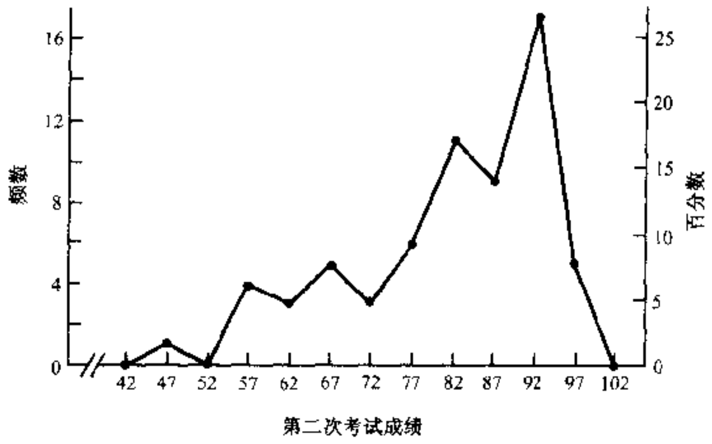


图 5-25

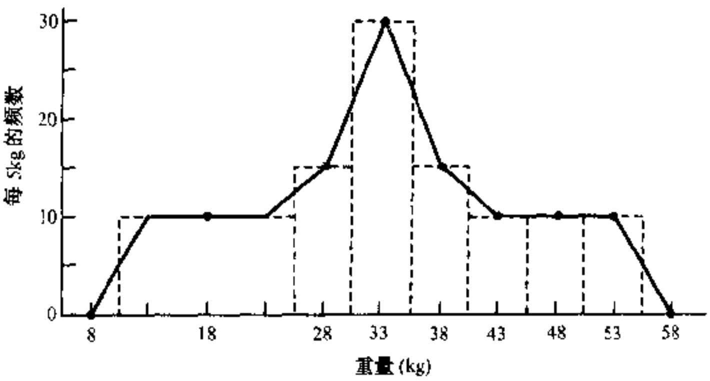


图 5-26

解 ☞　不能. 百分数折线图是封闭的平面图形, 因此对于有开端点组的分组数据不能构造折线图.

频数曲线、相对频数曲线和百分数曲线

5.21　图 5-27 是一些分布的理论频数曲线. 对每一条曲线, 从下列各项中选择可以描述其分布的项: 对称的, 正偏斜的, 负偏斜的, J 型的, 反 J 型的, 单峰的, 双峰的, 三峰的.

解 ☞　(a) 对称的,双峰的

(b) 对称的,单峰的

(c) 负偏斜的,单峰的

(d) 对称的,双峰的(也称为 **U 型分布**)

(e) 正偏斜的,反 J 型的

(f) 正偏斜的,双峰的

(g) 对称的(也称为**均匀分布、矩形分布**)

(h) 三峰的

(i) 负偏斜的,J 型的

茎叶表示法

5.22　将例 5.8 中的数据用**伸展的茎叶表示法**表示,仍然使用 1 位数的开始部分和叶,但是茎

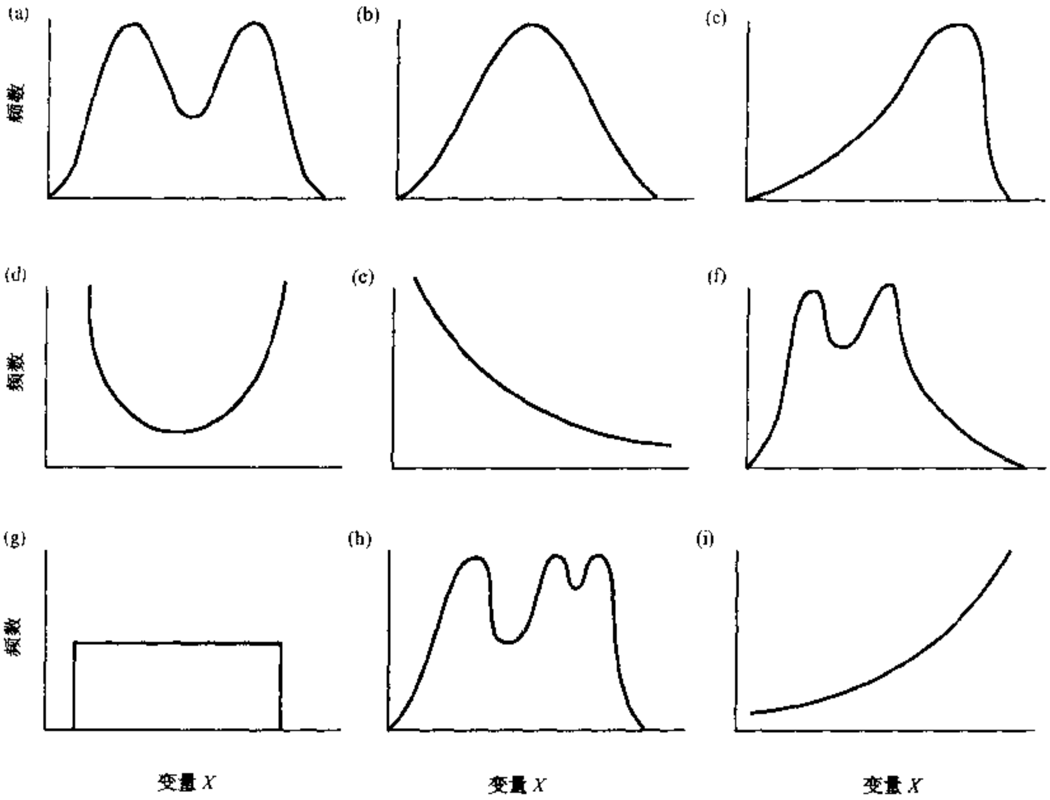


图 5-27

距为 5.

解 ☞　图 5-28 即为所要求的伸展的茎叶表示. 图 5-9 是简单的茎叶表示. 每一条茎有唯一的开始部分，茎距为 10. 而在图 5-28 中，茎距减为原来的一半，每条茎都伸展成两条茎距均为 5 的茎，它们的开始部分相同. 图 5-9 中的第一条茎是 1 | 5536976461，在这里被伸展成两条茎：1a | 341 和 1b | 5569766. 开始部分的右端字母 a 和 b 分别表示每一个开始部分的第一条茎和第二条茎. 简单茎叶表示的茎距可以是 0.1，1，10，100 等等，伸展的茎叶表示的茎距则是 0.05，0.5，5，50 等等.

1a	341	(3)
1b	5569766	(7)
2a	014124	(6)
2b	65957	(5)
3a	04113	(5)
3b	5799	(4)
4a	3400230	(7)
4b	655	(3)
		(40)

图 5-28

从图 5-9 看出，样本的频数分布呈矩形(见图 5-27(g))，但是在图 5-28 中它的形状变成双峰的(峰出现在开始部分 1b 和 4a). 这说明如果希望给出这些数据的分组分布，组距应为 5 或更小.

5.23　将例 5.8 中的数据用**压缩的茎叶表示法**表示，仍然使用 1 位数的开始部分和叶，但是茎距为 2.

解 ☞　图 5-29 即为所要求的压缩的茎叶表示，可以看出，图 5-9 中茎距为 10 的每一条原始茎，现在被分成("压缩成")5 条茎距均为 2 的茎. 它们的开始部分都与原始茎相同，但分别用字母 a，b，c，d 和 e 表示，以示区别. 因此，开始部分为 1a 的茎包括数 10 和 11，开始部分为 1b 的茎包括数 12 和 13，

1a	1	(1)
1b	3	(1)
1c	554	(3)
1d	6766	(4)
1e	9	(1)
2a	011	(3)
2b	2	(1)
2c	5445	(4)
2d	67	(2)
2e	9	(1)
3a	011	(3)
3b	3	(1)
3c	54	(2)
3d	7	(1)
3e	99	(2)
4a	000	(3)
4b	323	(3)
4c	545	(3)
4d	6	(1)
		(40)

图 5-29

开始部分为 1c 的茎包括数 14 和 15,开始部分为 1d 的茎包括数 16 和 17,开始部分为 1e 的茎包括数 18 和 19. 压缩的茎叶表示的茎距通常为 0.02,0.2,2,20 等等.

图 5-28 中的伸展的茎叶表示比图 5-9 中的简单茎叶表示有所进步,它展现了数据分布的双峰特征,并指出这些数据的分组分布的组距应为 5 或更小. 但是在压缩的茎叶表示中,数据太分散了,不能观察出分布的双峰特征. 因此这三个快速表示说明,这些数据的分组分布的组距应为 5 或更小,但是大于 2.

5.24 将附表 A.2 中全班同学的体重测量值(第 5 列)用**两侧简单茎叶表示法**(左侧为女生,右侧为男生)表示,开始部分使用两位数字,叶使用一位数字,茎距为 10.

解 ☞ 所要求的两侧简单茎叶表示见图 5-30. 当已知或怀疑同一个样本来自两个总体时,使用这种两侧表示法. 这种表示法可以快速的、非正式的比较两个总体. 对两个总体使用相同的开始部分,分别用向左侧延伸的叶和向右侧延伸的叶表示两个总体.

	女子		男子	
(2)	59	10		
(4)	1597	11	9	(1)
(7)	1143937	12	44	(2)
(3)	164	13	881	(3)
		14	82707394	(8)
		15	287622092181	(12)
		16	340900651	(9)
		17	2052360	(7)
		18	046	(3)
		19	014	(3)
(16)				(48)

图 5-30

累积分布图

5.25 对于表 4.42 中的分布,在一个坐标系中同时构造"小于式"累积频数曲线和"**小于式**"**累积百分数曲线**(或"**小于式**"**累积百分数折线图**).

解 ☞ 图 5-31 即为所要求的累积曲线,其构造方法见 5.11 节. 对于图中的每一个点,左边的纵

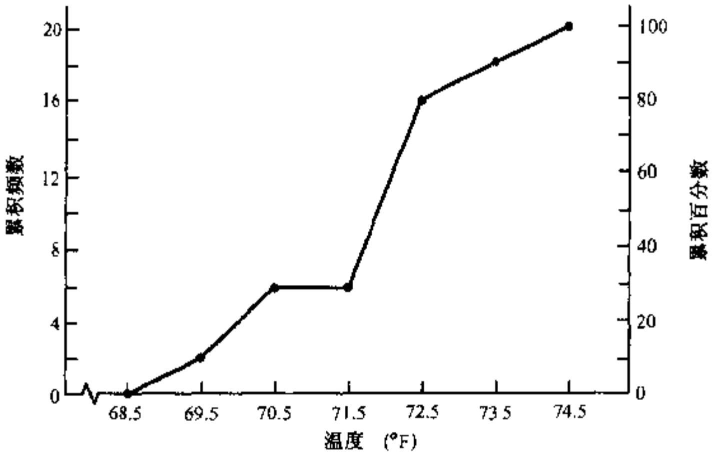


图 5-31

轴表示该点的"小于式"累积频数,右边的纵轴表示该点的"小于式"累积百分数.

5.26 为了进行遗传学研究,记录14只甲虫样本身体上的斑点数,表5.9是这些数据的频数分布. 请先将该频数分布转换成"小于式"累积频数分布和"小于等于式"累积频数分布(见4.9节). 然后在同一个坐标系中用图形表示这两个累积分布.

解 "小于式"累积频数分布见表5.10. 由于数据是离散比例水平的,所以累积频数说明有多少测量值小于测量类自身值(见习题4.9). 表5.11是"小于等于式"累积频数分布,每一类的累积频数是小于等于该类测量值的数据个数.

由于数据是离散比例水平，所以累积分布可以用两种图形表示. 它们可以用强调数据离散性质的图形表示，或者用假定它们是连续的图形表示（见习题5.9). 如果认为数据是连续的，那么能用累积曲线表示累积分布，见图5-32，该图中的两条累积曲线分别表示表5.10和表5.11中的累积分布. 对于两个相邻的测量类，两条累积曲线之间的水平虚线说明较大类的"小于式"累积频数等于较小类的"小于等于式"累积频数. 例如，有5个测量值小于3，同样有5个测量值小于等于2. 水平线也表明离散数据没有中间的值，所以两相邻测量值之间的"小于式"和"小于等于式"累积频数没有变化.

如果为了强调数据的离散性质,那么只画出图5-32中的水平线,见图5-33. 至今为止,这类图形没有统一的名称,本书称之为**累积分布的离散数据图形**.

表 5.9

斑点数 x_i	频数 f_i
1	2
2	3
3	4
4	3
5	2
$\sum$	14

表 5.10

斑点数	累积频数
小于1	0
小于2	2
小于3	5
小于4	9
小于5	12
小于6	14

表 5.11

斑点数	累积频数
小于等于0	0
小于等于1	2
小于等于2	5
小于等于3	9
小于等于4	12
小于等于5	14

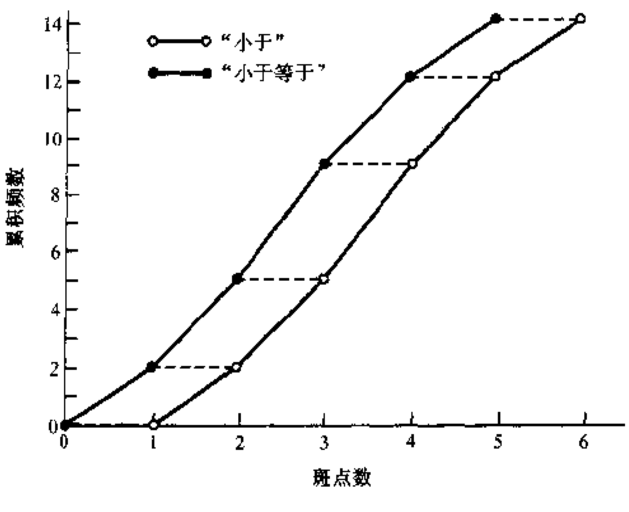


图5-32

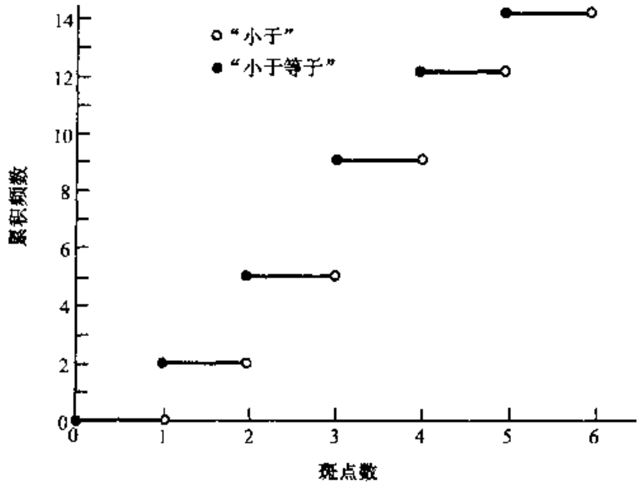


图 5-33

5.27 将表 5.9 中的频数分布转换成"大于等于式"累积频数分布和"大于式"累积频数分布(见 4.9 节). 然后在同一个坐标系中用图形表示这两个累积分布.

解 ☞ "大于等于式"累积频数分布见表 5.12. 由于数据是离散比例水平的,所以累积频数说明有多少测量值等于或大于某个给定的测量类. 表 5.13 是"大于式"累积频数分布,每一类的累积频数是大于该类测量值的数据个数.

由于数据是离散比例水平,所以可以用累积曲线或离散数据图形表示累积分布. 图 5-34 是这两个累积分布的累积曲线,图 5-35 是它们的离散数据图形. 图 5-34 中两相邻类的两条累积曲线之间的水平虚线说明较大类的"大于等于式"累积频数等于较小类的"大于式"累积频数,并且两相邻类之间的"大于式"和"大于等于式"累积频数没有变化. 与图 5-33 类似,数据的离散数据图形(图 5-35)只将图 5-34 中累积曲线之间的水平线画出.

表 5.12

斑点数	累积频数
大于等于 1	14
大于等于 2	12
大于等于 3	9
大于等于 4	5
大于等于 5	2
大于等于 6	0

表 5.13

斑点数	累积频数
大于 0	14
大于 1	12
大于 2	9
大于 3	5
大于 4	2
大于 5	0

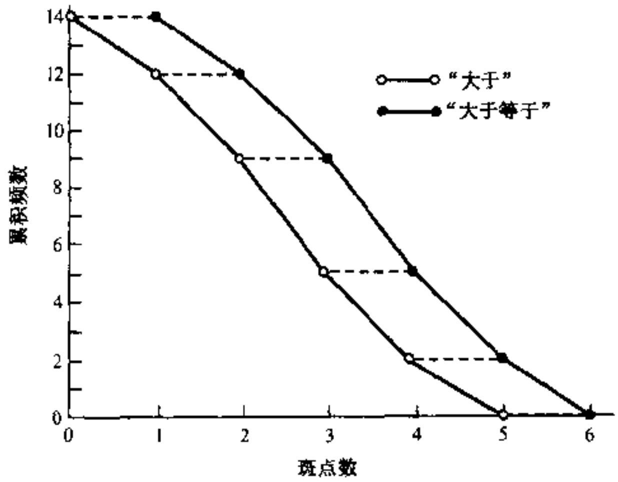


图 5-34

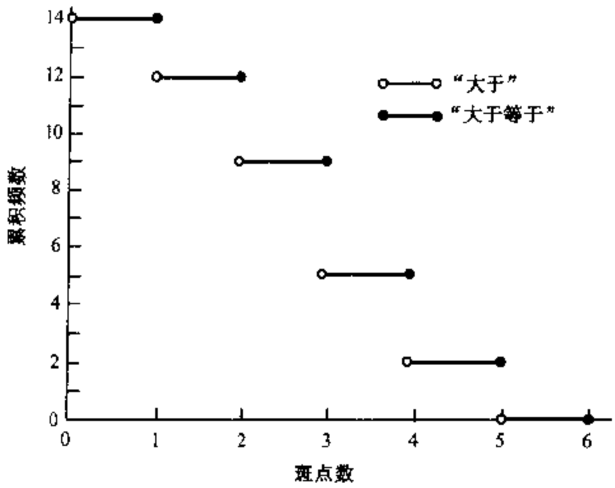


图 5-35

5.28 将表 4.12 中马拉松比赛成绩的分组"小于式"累积频数分布用"小于式"累积频数曲线表示.

解 ☞　所要求的累积频数曲线见图 5-36. 图中 Y 轴仍然表示累积频数,但 X 轴上标出的是组的上界.(对于这类分组数据,同样可以用 X 轴表示组的下限.)表示"小于式"累积频数的点处在组上界的上方,并用直线将这些点相连. 从图中看出,有 8 名运动员完成比赛的时间小于 136.5 min,所有运动员完成比赛的时间小于 145.5 min.

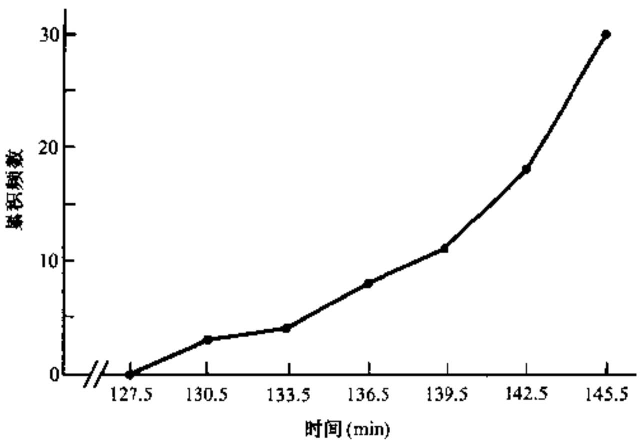


图 5-36

5.29 将表 4.22 中期末考试成绩的分组百分数分布转换成分组"小于式"累积百分数分布. 然后将这些离散数据看成"连续的",使用累积百分数曲线表示该分布,在每组的下限的上方描点. 最后从图中找出 50%的学生考试成绩比它低的分数.

解 ☞　所要求的"小于式"累积百分数分布见表 5.14,所要求的累积百分数曲线见图 5-37. 为了在图中**插值**(或**近似**)得到所要求的分数(50%的成绩"小于"它),第一步是过 Y 轴的 50%的刻度作一水平线延伸至累积曲线,该水平线与累积曲线相交于点 E. 点 E 的横坐标即为所要求的分数. 下一步是过点 E 作垂直线,与 X 轴相交. 由 X 轴的刻度可直接看出垂直线与 X 轴相交处是 84.5,这就是所求的分数的近似值.

这个分数的精确解可以通过相似三角形 ABC 和 ADE 计算出来. 这两个三角形的边长具有如下关系:

$$\frac{AD}{AB}=\frac{DE}{BC}$$

已知 AB=组距=5,BC=51.5625−34.3750=17.1875,DE=50−34.3750=15.6250,因此可以解出 AD:

$$\frac{AD}{5}=\frac{15.6250}{17.1875}$$

$$AD=5\times 0.9091=4.5455$$

因此,所求的分数是:80.0+4.5455=84.5455,或 84.5.

这个测量值(样本中有 50%的测量值"小于"它)称为样本的**中位数**. 中位数和其他的**集中趋势**、**平均值**和**位置**的度量将在第六章讨论.

表 5.14

第二次考试成绩	累积百分数
小于 45	0.0000
小于 50	1.5625
小于 55	1.5625
小于 60	7.8125
小于 65	12.5000
小于 70	20.3125
小于 75	25.0000
小于 80	34.3750
小于 85	51.5625
小于 90	65.6250
小于 95	92.1875
小于 100	100.0000

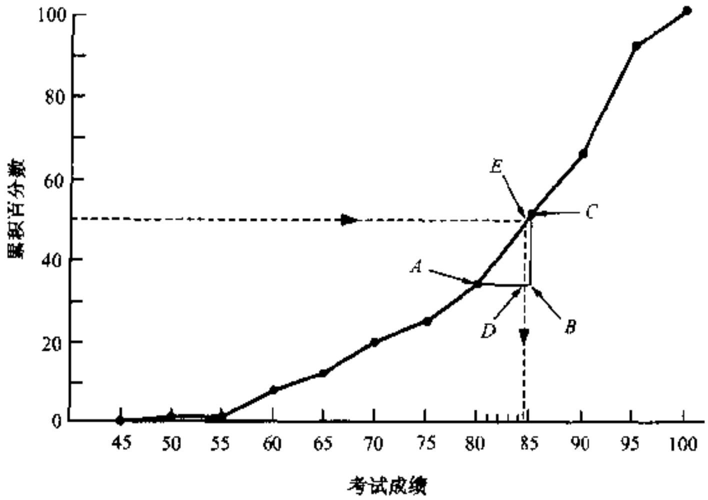


图 5-37

5.30 将表 4.31 中高尔夫选手所获奖金的分组"小于式"累积百分数分布用"小于式"累积百分数曲线表示,利用 X 轴上的测量尺度,在组的上界上方描出"小于式"百分数.

解 ☞ 所要求的累积百分数曲线见图 5-38. 注意,具有不等组距的分组分布的图形构造方法与等组距的分组分布的基本相同.

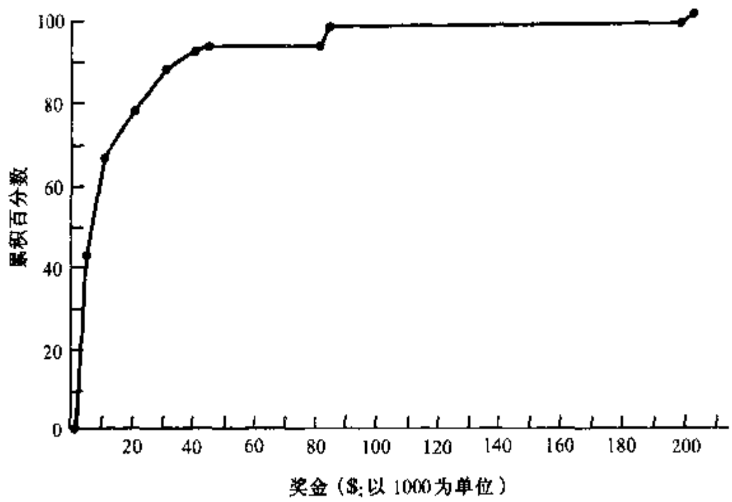


图 5-38

5.31 什么是**平滑曲线累积频数曲线**?

解 ☞ 在 5.7 节中我们指出,当样本容量 n 接近总体容量 N 时,频数折线图会变成一平滑曲线,称为频数曲线. 类似地,当样本容量接近总体容量时,样本累积频数曲线变成一**平滑曲线累积频数曲线**(累积相对频数曲线变成一**平滑曲线累积相对频数曲线**,百分数累积曲线变成一**平滑曲线百分数累积曲线**).

补充习题

条形图

5.32 下届市长选举中有三位候选人:A,B和C. 为了了解候选人受欢迎程度,抽取了100名选举人进行民意测验. 图5-39为候选人C构造的民意测验结果的频数条形图. 请问每位候选人的支持者分别为多少? 为什么这个条形图不是测验结果的真实表示,具有欺骗性?

答案 民意测验的结果非常接近:候选人A的得票为31,候选人B的得票为33,候选人C的得票为36. 但是从条形图分析,似乎候选人C获得了胜利,这里有三个迷惑的地方:不从Y轴的零点开始,而是从30开始,从而夸大了矩形间的差别;横轴为纵轴的75%;用来表示候选人C的矩形的底宽是其它矩形的两倍,使得该矩形的面积非常大. 所以说这个条形图具有欺骗性.

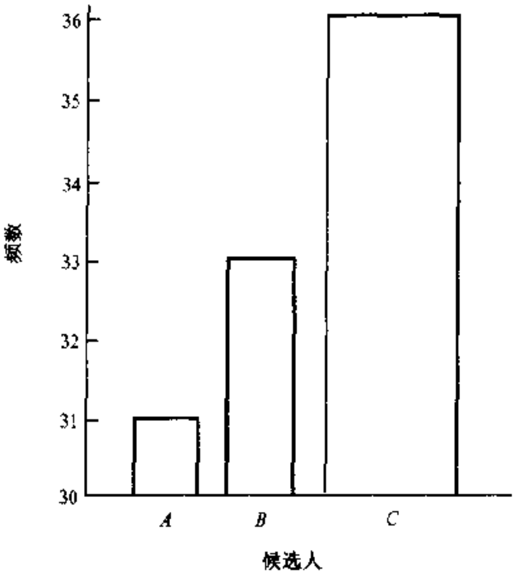


图5-39

5.33 广口瓶中有黄色(Y)、红色(R)、蓝色(B)和白色(W)四种球,共200个. 图5-40是它们的条形图,其中每个矩形表示一种颜色类,右端的纵轴表示相对频数尺度. 根据所给信息,请确定瓶中每种颜色球的数量,并给出左侧纵轴频数尺度上的5条刻度线的数值.

答案 20 Y,20 R,60 B,100 W;刻度值为:20,40,60,80,100

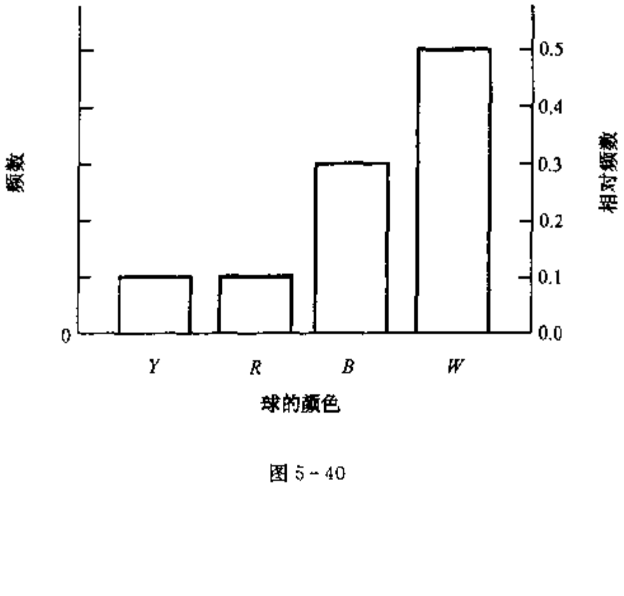


图5-40

直方图

5.34 对于下列测量变量,指出其测量水平,以及适合的柱状图类型:(a) 产仔数量(如表 4.13),(b) 用 5 个数字测量美国人对移民的态度,1=不欢迎,……,5=非常欢迎.(c) 体温(°F),(d) 蜗牛外壳的直径(mm)

答案 (a) 离散比例水平;条形图,杆图或直方图,(b) 次序水平;条形图,(c) 间隔水平;直方图,(d) 连续比例水平;直方图

5.35 对于下列测量变量,指出其测量水平,以及适合的柱状图类型:(a) 每分钟录入的字数,(b) 在几个指定区域汽油的价格(美元/加仑),(c) 维生素的类型,(d) 一队卡车中每辆卡车每年运行的公里数

答案 (a) 离散比例水平;条形图,杆图或直方图,(b) 离散比例水平;条形图,杆图或直方图,(c) 名义水平;直方图,(d) 连续比例水平;直方图

5.36 将表 4.38 中统计专业班级的所有同学(男生和女生)的身高测量值用频数直方图表示,在 X 轴标出测量尺度所使用的部分.

答案 图 5-41 即为所要求的直方图.

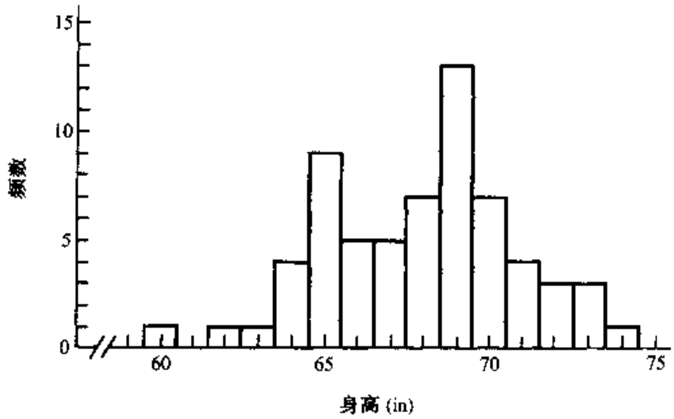


图 5-41

5.37 对于图 5-41 中的数据,构造男生和女生的**成分部分频数直方图**,仍然在 X 轴上标出测量尺度所使用的部分.

答案 图 5-42 即为所要求的直方图. 图 5-41 显示数据是双峰的,这通常说明样本来自两个不同的总体. 图 5-42 确认了这种解释,表明这两个总体是男生和女生.

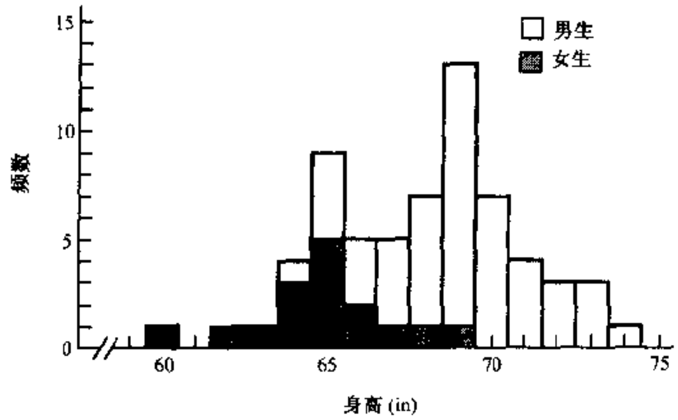


图 5-42

折线图

5.38 为表 5.5 中的数据构造折线图,同时标出频数和相对频数. 不用给出相关的直方图,而是直接在坐标系中描点.

答案 图 5-43 即为所要求的折线图. 注意到,这个单峰的、正偏斜的折线图在零频数类 105°F 下降到 X 轴.

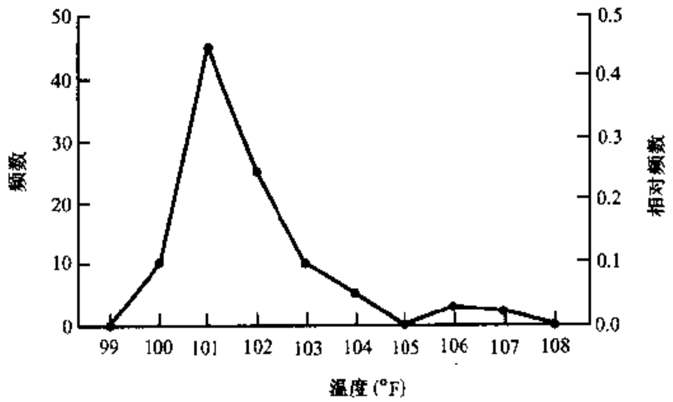


图 5-43

5.39 一个录像带出租店的雇员将上星期的出租情况进行了汇总,图 5-44 是每位顾客所租录像带总数的折线图. 请问:(a) 上星期有多少顾客租了录像带? (b) 租两盒录像带的顾客所占的比例是多少?

答案 (a) 100,(b) 0.5

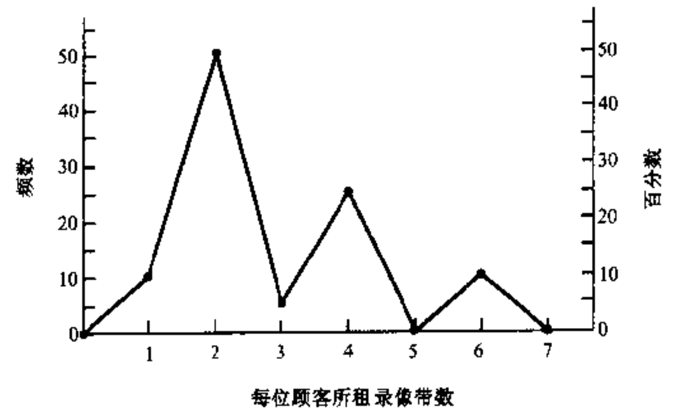


图 5-44

5.40 根据图 5-44 给出的信息,回答下列问题:(a) 租 4 盒或更少的录像带的顾客所占的比例为多少? (b) 有多少顾客租了 3 盒或更多的录像带?

答案 (a) 90%,(b) 40

5.41 为了进行质量控制,一家水果进口公司测量了一些甜瓜样本的重量. 将测量结果分组(采用等组距)后用折线图表示,见图 5-45,测量尺度见 X 轴. 请问:(a) 样本的容量是多少? (b) 数据被分成了几组?

答案 (a) 40,(b) 7

5.42 根据图 5-45 给出的信息,回答下列问题:(a) 这 7 个组的组中值分别是多少? (b) 这些组的组界是什么?

答案 (a) 组中值为:1.3,1.6,1.9,2.2,2.5,2.8,3.1,(b) 组界为:(1.15 到 1.45),(1.45 到 1.75),(1.75 到 2.05),(2.05 到 2.35),(2.35 到 2.65),(2.65 到 2.95),(2.95 到 3.25).

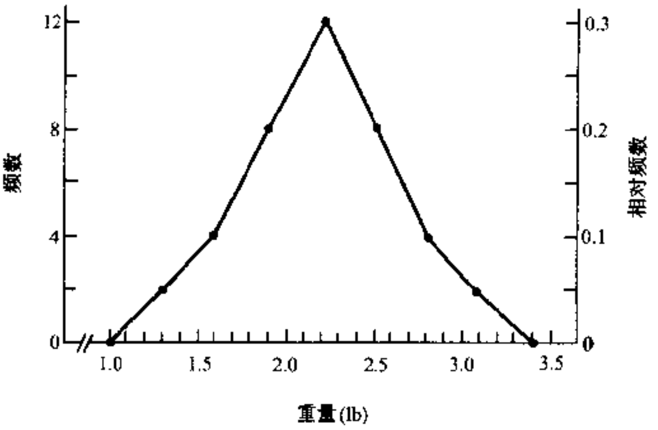


图 5－45

茎叶表示法

5.43 请将下列数据集用简单茎叶表示法表示，开始部分和叶均使用一位数字，茎距为 1：

3.792，7.300，1.419，8.333，3.212，2.513，2.937，5.312，4.821，1.694，2.100，7.902，9.111，2.321，2.119，6.199，8.774，2.572，3.999，3.192，5.988，2.412，4.911，6.900，7.297，2.633，4.431，5.255，6.591，4.497，6.511，2.617.

答案 图 5－46 即为所要求的简单茎叶表示图，原始数据的小数点右边都有三位数字，但是在该图中只记录了小数点右边第一位数字.

1	46	(2)
2	591315466	(9)
3	7291	(4)
4	8944	(4)
5	392	(3)
6	1955	(4)
7	392	(3)
8	37	(2)
9	1	(1)
		(32)

图 5－46

5.44 请将习题 5.43 中的数据集用茎叶表示法表示，开始部分使用一位数字，叶使用三位数字，茎距为 1.

答案 图 5－47 即为所要求的茎叶表示图. 如该图所示，如果叶中的数字超过一位，那么相邻的叶用逗号隔开.

1	419,694	(2)
2	513,937,100,321,119,572,412,633,617	(9)
3	792,212,999,192	(4)
4	821,911,431,497	(4)
5	312,988,255	(3)
6	199,900,591,511	(4)
7	300,902,297	(3)
8	333,774	(2)
9	111	(1)
		(32)

图 5－47

累积分布图

5.45 在一次国际田径运动会中，为了确定参加男子 200 米决赛的 8 名运动员，举行了两组半决赛. 半决赛成绩的累积曲线见图 5－48；第 1 组的"大于等于式"累积频数曲线，其中频数的累积是从测量类的蕴含范围的下界开始；第 2 组是"小于式"累积频数曲线，其中频数的累积是从测量类的蕴含范围的上界开始. 请问：(a) 有多少运动员参加了半决赛？(b) 有多少运动员的比赛成绩为 21.0 秒？

答案 (a) 第 1 组 8 个，第 2 组 7 个，(b)0

5.46 根据图 5－48 给出的信息，回答下列问题：(a) 前 8 名运动员的比赛成绩是多少？(b) 两组半决赛中最慢的运动员的比赛成绩是多少？

答案 (a) [(第 1 组)20.8 秒]，[(第 2 组)20.9 秒]，[(第 1 组)21.1 秒，21.1 秒]，[(第 2 组)21.1 秒]，[(第 2 组)21.2 秒，21.2 秒，21.2 秒]，(b) [(第 1 组)21.5 秒]，[(第 2 组)21.4 秒]

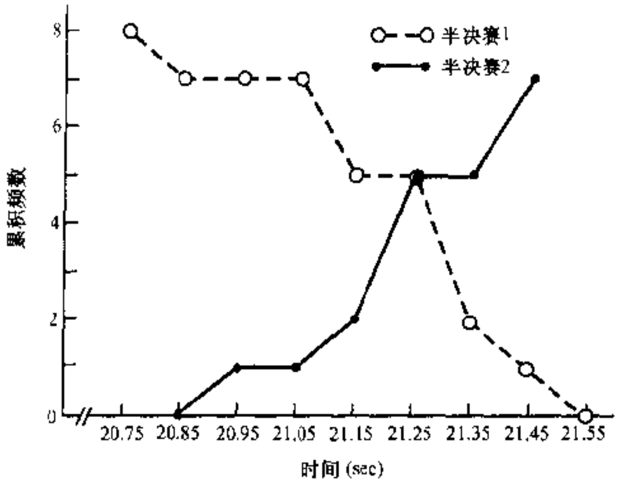


图 5 - 48

5.47　将表 4.43 中产仔数量的“大于等于式”累积百分数分布用离散数据图形表示.

答案　所要求的离散数据图形如图 5 - 49 所示.

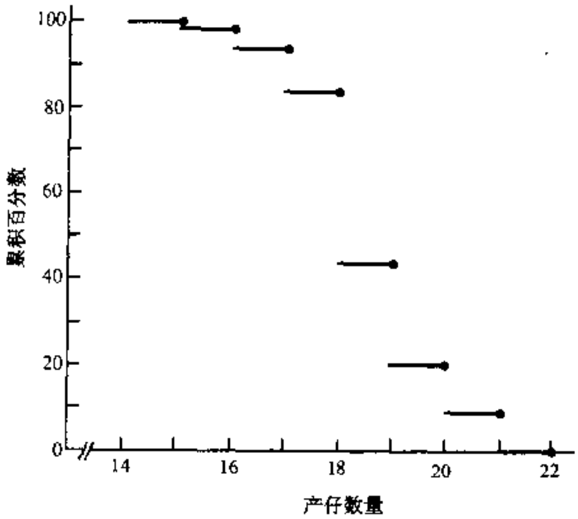


图 5 - 49

5.48　对于表 4.44 中学生体重的“小于式”累积频数分布,在同一个坐标系中分别构造女生和男生的“小于式”累积频数曲线,在 X 轴标出上界.

答案　所要求的累积曲线如图 5 - 50 所示.

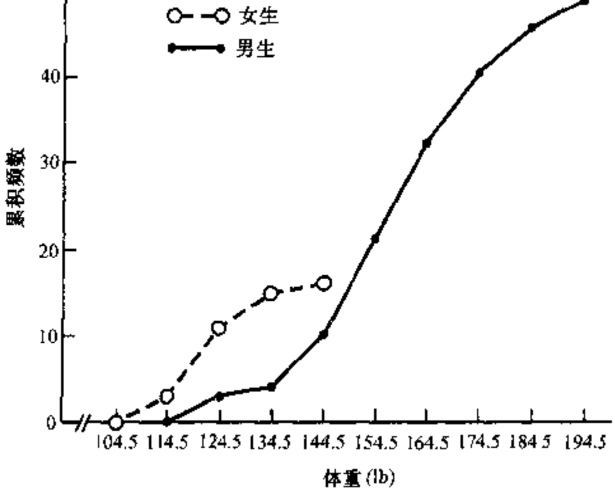


图 5 - 50

第六章　描述性统计：集中趋势、平均值和位置的度量

6.1　集中趋势、平均值和位置的度量

上两章介绍了描述统计学的两个基本部分：统计表（第四章）和统计图（第五章）. 这一章和下一章将介绍描述统计学的另一个方面，统计资料的**描述性度量**的计算，也就是统计资料的度量，即用一些数值，特别是单个的数字，表示统计资料的特征. 这一章中我们介绍描述**集中趋势**、**平均值**和**位置**的特征数，第七章中将介绍描述**离散性**（数据在分布中的散布）的度量. 对于其他的度量则在需要用到时再作介绍.

描述性统计度量有两个作用：首先，它们为受过统计训练的人提供数据分布的初步印象. 其次，由于大多数样本的描述性度量（即统计量）常用于估计相应的总体度量（即参数），所以描述性统计度量是推断统计学的主要成分，是参数估计和假设检验的基础（见 3.6 节）.

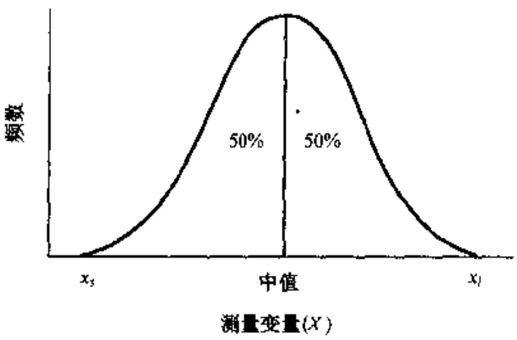


图 6－1

图 6－1 是一个对称的、单峰的频数曲线，我们考察它的某些特征. 在该曲线中，最大频数发生在最小测量值 x_s 与最大测量值 x_l 之间的中间位置. 这种测量值在分布中心位置的聚集现象称为分布的**集中趋势**，这是许多数据类型都具有的特征，一个分布的中心特征的统计度量称为**集中趋势的度量**.

数据集中的**平均值**是最典型的、最常用的、最具代表性的度量. 由于测量值常常集中在分布的中心，所以集中趋势的各种不同的度量一般也称为**平均值的度量**.

位置的度量反映一个分布特征关于测量尺度所在的位置. 例如，图 6－1 中有三种不同的位置度量：最小值 x_s、最大值 x_l 和中位数（在它的左、右各有 50%数据的分界点，见习题 5.29）. 由于集中趋势和平均值的度量说明了分布中心位置的统计资料的特征，所以有的统计书上也将这些度量统称为位置的度量.

6.2　算术平均数

算术平均数可以用下列诸式定义：

$$\bar{x}=\frac{\sum_{i=1}^{n} x_i}{n} \tag{6.1}$$

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i \tag{6.2}$$

$$\bar{x}=\frac{\sum x_i}{n} \tag{6.3}$$

$$\mu=\frac{\sum_{i=1}^{N} x_i}{N} \tag{6.4}$$

$$\mu=\frac{1}{N}\sum_{i=1}^{N} x_i \tag{6.5}$$

$$\mu=\frac{\sum x_i}{N} \tag{6.6}$$

式(6.1)和(6.4)已经在3.4节中介绍了.式(6.1)计算的是样本$x_1,\cdots,x_n$的算术平均数,即样本均值$\bar{x}$,其中n表示样本容量.式(6.4)计算的则是总体$x_1,\cdots,x_N$的算术平均数,即总体均值μ,其中N表示总体容量.式(6.2)与(6.1)相同,式(6.5)与(6.4)相同,不同的是用和乘以$1/n$或$1/N$,而不是用和除以n或N(通常用准确值做除法,而不是对舍入运算后的值做乘法).

如果求和下标(见1.22节)没有特别指明,则意味着数据集中的所有数据都参与求和.因此,式(6.3)等价于式(6.1)和(6.2),式(6.6)等价于式(6.4)和(6.5).

算术平均数是最常用的集中趋势、平均值和位置的度量.所以当提到"平均数"时,例如平均击球数、平均价格、年平均降雨量等等,通常都理解为算术平均数.然而,这种解释并不一定正确,因为还有其他的度量也称为均值和平均数.算术平均数无疑是推断统计学中最重要的一种度量,样本均值$\bar{x}$被认为是总体均值μ的最可信、最有效的估计.算术平均数一般只用于间隔和比例型测量变量,但是对于次序型数据,它仍然适用.

例 6.1　计算下列样本的算术平均数:(a) $x_1=1$ g,$x_2=3$ g,$x_3=2$ g,$x_4=7$ g,$x_5=5$ g,$x_6=4$ g,$x_7=2$ g,(b) 1 g,3 g,2 g,7 g,5 g,4 g,200 g

解　(a) $\bar{x}=\dfrac{\sum x_i}{n}=\dfrac{1\text{ g}+3\text{ g}+2\text{ g}+7\text{ g}+5\text{ g}+4\text{ g}+2\text{ g}}{7}=\dfrac{24\text{ g}}{7}=3.4\text{ g}$

(b) $\bar{x}=\dfrac{\sum x_i}{n}=\dfrac{1\text{ g}+3\text{ g}+2\text{ g}+7\text{ g}+5\text{ g}+4\text{ g}+200\text{ g}}{7}=\dfrac{222\text{ g}}{7}=31.7\text{ g}$

注:(b)中没有用x_i给出数据,这里是假定测量值以$x_1,x_2,\cdots,x_n$的顺序给出.本书中没有给出求和下标处都遵从此假定.从这些计算可知,算术平均数与测量值具有相同的单位.如果一个数据与所在数据集中的其余数据有很大的不同,则称为**极端值或异常值**.从(a)和(b)的计算结果可以看出,算术平均数受数据中极端值的影响很大,例如在(a)、(b)中当x_7从2变成为200时,平均值从3.4 g变到31.7 g.数据集中极端值的出现通常意味着某种程序上的错误或试验失败.但是,极端值也有可能是真实数据,这时说明某些外部变量(见3.10节)的影响作用.

6.3　算术平均数的舍入准则

在计算算术平均数那样的描述性度量时,有两个相关但不同的舍入问题:(1)在计算的不同步骤中什么时候应进行舍入运算,怎样进行舍入运算,(2)在最后的答案中应保留几位数字.

本书中的大多数描述性度量都是用公式定义的,且度量值**精确**等于按指定公式计算的最后结果.例如,$\bar{x}$精确等于数据之和除以样本容量,但这只在理论上成立,实际中通常不可能得到精确值,如果这样,在计算的每一步都要保留无限多位数字,而这是不可能的.为了得到与精确值尽量接近的结果,**在计算过程中应尽可能少的进行舍入运算**.对于计算过程中保留几位数字的问题并没有统一的准则,但是有的书上建议至少保留6位数字.

如果描述性度量值还需用于更进一步的计算中,那么其计算结果应保留多位数字.如果描述性度量值是一个最后的计算结果,那么这时应保留几位数字?这个问题有许多准则可供参考.对于算术平均数最常用的准则是:

如果所有数据的精密度水平(见2.15节)相同,那么平均数应使用下一个精密度水平表示.

我们在例6.1中使用的也是这一准则.在例6.1(a)中,所有的数据都是精确到个位,所以算术平均数精确到于分位.使用12位数字的计算器得到$\bar{x}$为3.42857142857,精确到于分位,得到$\bar{x}=3.4$ g.

另外一些常用的准则是:算术平均数用与数据相同的精密度水平或下一个精密度水平表示;算术平均数与**标准差**(见第七章)及**均值的标准误差**(见第十三章)保留相同位数的数字.对于其他的重要准则在用到时再作介绍.如果没有指明舍入准则,那么使用最常用的准则.

最后要指出的是,本书所有的计算都是使用最多显示12位数字的计算器,**其小数点位数**最多为11位,并且该计算器的计算过程中的数值保留15位数字.因此,如果你使用的是更少位数的计算器,那么得到的结果可能与本书结果稍有不同.

例 6.2　计算下列样本的算术平均数:(a) -3℃,$+2$℃,-1℃,-4℃,-6℃,-5℃,(b)

2.002 g, 3.7 g, 2.963 g, 3.5041 g, 2.737 g, 1.99999 g.

解 ☞ (a)中数据的精密度水平相同，所以使用最常用的舍入准则. 而(b)中数据的精密度水平不全相同，这时我们要使用代数运算的基本舍入准则(见 2.15 节).

(a) $\bar{x}=\dfrac{\sum x_i}{n}=\dfrac{(-3℃)+2℃+(-1℃)+4℃+(-6℃)+(-5℃)}{6}=\dfrac{-9℃}{6}=-1.5℃$

(b) $\bar{x}=\dfrac{\sum x_i}{n}=\dfrac{2.002\text{ g}+3.7\text{ g}+2.963\text{ g}+3.5041\text{ g}+2.737\text{ g}+1.99999\text{ g}}{6}$

$=\dfrac{16.9\text{ g}}{6}=2.81666666667\text{ g}$，或 2.82 g

6.4 与算术平均数的离差和分布重心

总体中测量值与总体算术平均数的差(或距离)，$x_i-\mu$，称为**测量值与总体均值的离差**(或简称为**均值离差**). 同样地，样本中测量值与样本算术平均数的差，$x_i-\bar{x}$，称为**测量值与样本均值的离差**. 在频数直方图中，均值左边的测量值(比均值小)的离差是负的，称为**负离差**；而均值右边的测量值(比均值大)的离差是正的，称为**正离差**. 算术平均数的一个重要性质是：无论是总体还是样本，其正离差和负离差能完全抵消. 下面给出这一性质：对于总体，$\sum_{i=1}^{N}(x_i-\mu)=0$ 的数学证明.

对于任何给定的总体，其均值 μ 可以认为是一个常数，所以(见习题 1.45)

$$\sum_{i=1}^{N}(x_i-\mu)=\sum_{i=1}^{N}[x_i+(-\mu)]=\sum_{i=1}^{N}x_i+\sum_{i=1}^{N}-\mu=\sum_{i=1}^{N}x_i-N\mu$$

用 $\dfrac{\sum_{i=1}^{N}x_i}{N}$ 代替 μ

$$\sum_{i=1}^{N}(x_i-\mu)=\sum_{i=1}^{N}x_i-N\left(\frac{\sum_{i=1}^{N}x_i}{N}\right)=\sum_{i=1}^{N}x_i-\sum_{i=1}^{N}x_i=0$$

同样的，对于样本可以证明，$\sum_{i=1}(x_i-\bar{x})=0$.

由于对任意数据集都有，算术平均数的离差和等于零，所以正离差之和总是等于负离差之和. 考虑数据的频数直方图，如果将离差看成是距离，那么，与均值的正距离的和等于负距离的和. 如果频数直方图是用单一材料制成的，那么它将在水平轴(X 轴)上算术平均数那点平衡. 所以算术平均数也称为分布**重心**的度量. 因为测量值聚集在分布的中心附近，所以分布重心的度量也是一种集中趋势和中心位置的度量(见 6.1 节).

6.5 平均值的一种度量——算术平均数

如 6.1 节中所述，一个数据集的平均值是最典型的、最常用的、最具代表性的度量. 下面的推导可以说明：算术平均数正是这样一种度量，所以它是**平均值的一种度量**.

由于

$$\mu=\frac{\sum_{i=1}^{N}x_i}{N}\quad 和\quad \bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$$

所以

$$N\mu=\sum_{i=1}^{N}x_i \tag{6.7}$$

$$n\bar{x}=\sum_{i=1}^{n}x_i \tag{6.8}$$

因此，如果数据集中的所有测量值都用它们的算术平均数代替，那么测量值的总和将保持不变．只有算术平均数才具有这个性质，因此从这个意义上来说，算术平均数是数据集的最具代表性的值．

6.6　由未分组频数分布计算算术平均数

如果样本已整理成表 5.1 的频数分布，那么其算术平均数也可以直接由公式(6.1)计算得到：

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{(1.2+1.2+1.3+\cdots+1.7+1.8+1.8)\text{cm}}{50}=1.50\text{ cm}$$

或者直接利用频数分布，得到一种更简单的方法计算样本的算术平均数．它的计算公式如下：

$$\bar{x}=\frac{\sum_{i=1}^{k}f_i x_i}{\sum_{i=1}^{k}f_i} \tag{6.9}$$

$$=\frac{\sum_{i=1}^{k}f_i x_i}{n} \tag{6.10}$$

其中，x_i 表示变量 X 的第 i 类，f_i 表示第 i 类的频数(见例 4.3)．

类似地，由未分组频数分布计算总体均值 μ 的公式为：

$$\mu=\frac{\sum_{i=1}^{k}f_i x_i}{N} \tag{6.11}$$

例 6.3　用(6.10)式计算表 5.1 中样本的均值．

解　为了使用(6.10)式，我们在表 5.1 中添加一列：$f_i x_i$，形成表 6.1，并在最后给出了计算结果．

注：由该数据的频数直方图(图 5-3)，可以看出 1.50 cm 恰好是这个单峰对称分布的重心平衡点，经过这点的垂直线将直方图分成两个相等的部分．我们将在例 6.13 中计算这个数据的中位数，可以看到它与算术平均数相等．

表 6.1

长度(cm) x_i	频数 f_i	$f_i x_i$(cm)
1.2	2	2.4
1.3	7	9.1
1.4	10	14.0
1.5	12	18.0
1.6	10	16.0
1.7	7	11.9
1.8	2	3.6
$\sum$	50	75.0 cm

$\bar{x}=\frac{\sum f_i x_i}{n}=\frac{75.0\text{ cm}}{50}=1.50\text{ cm}$

6.7　由分组频数分布计算近似算术平均数

由分组频数分布(见 4.4 节)计算的算术平均数是由数据直接计算的准确值的一个近似，因此称之为**近似算术平均数**．由分组频数分布计算算术平均数的前提假设是组中的所有测量值都等于组中值 m_i．计算总体的近似算术平均数的公式为

$$\mu\approx\frac{\sum_{i=1}^{k}f_i m_i}{N} \tag{6.12}$$

样本的近似算术平均数的计算公式为

$$\bar{x}\approx\frac{\sum_{i=1}^{k}f_i m_i}{n} \tag{6.13}$$

由于组中值是组限的算术平均数(见 4.4 节)，所以组中的所有测量值都等于组中值 m_i 这一前提假设是不合理的．在前面我们两次用到这一假设：习题 4.24(e) 中估计总的汽车销售量和 5.6 节中根据分组数据构造折线图．

例 6.4 根据表 4.3 中的未分组频数分布计算马拉松比赛中前 30 名男运动员成绩的**精确的算术平均数**. 然后再根据表 4.4 中的分组频数分布计算它们的近似算术平均数.

解 ☞ 表 6.2 给出了利用(6.11)式计算该数据的精确(或真实)算术平均数的过程及结果,表 6.3 给出的则是利用(6.12)式计算近似算术平均数的过程及结果.

表 6.2

时间(min) x_i	频数 f_i	f_ix_i(min)
129	1	129
130	2	260
131	0	0
132	0	0
133	1	133
134	1	134
135	1	135
136	2	272
137	0	0
138	3	414
139	0	0
140	0	0
141	3	423
142	4	568
143	5	715
144	2	288
145	5	725
$\sum$	30	4,196 min

$$\mu = \frac{\sum f_i x_i}{N} = \frac{4,196 \text{ min}}{30} = 139.9 \text{ min}$$

表 6.3

时间(min)	组中值 m_i	频数 f_i	f_im_i(min)
128—130	129	3	387
131—133	132	1	132
134—136	135	4	540
137—139	138	3	414
140—142	141	7	987
143—145	144	12	1,728
$\sum$		30	4,188 min

$$\mu \approx \frac{\sum f_i m_i}{N} = \frac{4,188 \text{ min}}{30} = 139.6 \text{ min}$$

注:从统计意义上来说,近似测量值的准确度比真实测量值的更低(见 2.14 节),所以只有在不可能计算真实测量值时才计算它. 从这个问题可以看出近似总体均值(139.6 min)低估了真实总体均值(139.9 min),这可能会严重影响到涉及均值的进一步计算. 在后面的例 6.14 中我们还将计算该数据的**中位数的精确值**和**近似值**,并可得出这样的结论,对于具有负偏态的分布(见图 5-6),中位数位于算术平均数的右边.

6.8 由编码数据计算算术平均数

如果数据集由很大或很小的数组成,并且没有计算机可以使用而需要计算如算术平均数那样的统计度量时,这时则常使用**编码公式**

$$c_i = a + bx_i \tag{6.14}$$

将数据**转换**成更简单的数后再计算算术平均数,其中 x_i 是变量 X 的第 i 个测量值,a 和 b 是常数,c_i 称为第 i 个测量值的**转换值**(或**编码值**).

当 $a \neq 0, b=1$ 时,每一个测量值都加上了一个常数,这时测量尺度的原点发生了平移,这种编码过程称为**数据的平移**.

当 $a=0, b>0$ 并且 $b \neq 1$ 时,每一个测量值都乘以了一个常数 b. 若 $b>1$,则是**测量尺度的扩展**,若 $0<b<1$,则是**测量尺度的压缩**. 这三种编码形式(**扩展**、**压缩**和**平移**)都是数据的**线性变换**.

编码值 c_i 的算术平均数可以用下面的公式计算:

$$\bar{c} = \frac{\sum_{i=1}^{n} c_i}{n} \tag{6.15}$$

原始数据的算术平均数可以用下面的**样本算术平均数的解码公式**计算得到:

$$\bar{x} = \frac{1}{b}(\bar{c} - a) \tag{6.16}$$

例 6.5　首先利用式(6.1)直接计算下列长度测量值(以 cm 为单位)的算术平均数:492,493,495,496,498,500. 然后根据式(6.14)、(6.15)和(6.16),并设 $a=-490$,$b=1$,计算这些数据的算术平均数.

解 ☞　$\bar{x}$ 的直接计算和使用编码、解码公式的计算过程和结果见表 6.4.

表 6.4

长度(cm) x_i	$c_i=-490\text{ cm}+x_i\text{ cm}$
492	2
493	3
495	5
496	6
498	8
500	10
$\sum$ 2,974 cm	34 cm

$$\bar{x}=\frac{\sum x_i}{n}=\frac{2,974\text{ cm}}{6}=495.7\text{ cm}$$

$$\bar{c}=\frac{\sum c_i}{n}=\frac{34\text{ cm}}{6}=5.66667\text{ cm}$$

$$\bar{x}=\frac{1}{b}(\bar{c}-a)=\frac{1}{1}[5.66667\text{ cm}-(-490\text{ cm})]=495.7\text{ cm}$$

6.9　加权平均

在 6.2 节中的算术平均数的计算公式是先将样本(或总体)中所有的值相加,再除以样本(或总体)中数据的个数. 这个公式中所有的数据都被认为具有**同等重要性**,所以在计算均值时使用的是**相等权重**. 但是,如果数据的不同类型对均值的贡献有所不同,那么在计算平均数时就应对每一种类型的数据赋予与其重要性成比例的**权重**. 这样计算的平均数就称为**加权平均数**(或**加权算术平均数**). 样本的加权平均数的计算公式为

$$\bar{x}_w=\frac{\sum_{i=1}^{k}w_ix_i}{\sum_{i=1}^{k}w_i} \tag{6.17}$$

其中 $\bar{x}_w$ 表示样本加权平均数,x_i 是变量 X 的第 i 个测量值,w_i 是第 i 个测量值的权重,k 是测量类的个数. 总体的加权平均数(μ_w)的计算公式为

$$\mu_w=\frac{\sum_{i=1}^{k}w_ix_i}{\sum_{i=1}^{k}w_i} \tag{6.18}$$

例 6.6　一玩具制造商有 50 名职工,其中 15 名职工的每小时工资为 \$5.25,25 名职工的每小时工资为 \$5.75,10 名职工的每小时工资为 \$6.30. 请利用(6.18)式计算这 50 名职工的平均每小时工资.

解 ☞　这个问题中的 x_i 是每小时工资,每一种工资水平的相对重要性 w_i,就是这个工资水平的职工人数 f_i. 因此,这个总体的加权平均数为

$$\mu_w=\frac{\sum_{i=1}^{k}w_ix_i}{\sum_{i=1}^{k}w_i}=\frac{\sum_{i=1}^{3}f_ix_i}{\sum_{i=1}^{3}f_i}=\frac{(15\times\$5.25)+(25\times\$5.75)+(10\times\$6.30)}{15+25+10}$$

$$=\frac{\$285.50}{50}=\$5.71$$

注:从这个例子可以看出,由未分组频数分布计算算术平均数的公式(见 6.6 节)是加权平均值计算公式的特殊情形,其中 $w_i=f_i$. 类似的,我们也可以将计算算术平均值的基本公式(见 6.2 节)看成加权平均值计算公式的一种特殊情形,其中 $w_i=1$,对于样本 $k=n$,对于总体 $k=N$. 由于 $w_i=1$,所以也将这些基本算术平均数称为**无权重算术平均数**或**简单算术平均数**. 近似算术平均数(见 6.7 节)也是加权平均的一种特殊情形,其中 $w_i=f_i$,$x_i=m_i$,k 是分组分布的组数.

6.10 总平均

将几个不同样本的算术平均数用适当的方法相结合的结果称为**总平均数**. 总平均数的计算公式也是加权平均[(6.17)式]的一种:

$$总平均数=\bar{x}_w=\frac{\sum_{i=1}^{k}n_i\bar{x}_i}{\sum_{i=1}^{k}n_i} \tag{6.19}$$

其中权重 w_i 是样本容量 n_i,x_i 是样本的均值 $\bar{x}_i$,k 是要进行合并的样本个数. 由(6.8)式知,$n\bar{x}=\sum_{i=1}^{n}x_i$,因此总平均数公式中的分子 $\sum_{i=1}^{k}n_i\bar{x}_i$ 就是所有样本中的所有数据的和. 而分母是样本容量的和,所以总平均数实际上就是所有数据的和除以数据的个数.

例 6.7 要研究一种新型降压药物的药效,从三个医院各选一组女性患者进行研究,分别测量采用这种药物治疗前后患者的每分钟心跳次数. 采用这种药物治疗之前的测量结果为:医院 1,$n_1=30$ 个病人,$\bar{x}_1=76.2$ 次/分钟;医院 2,$n_2=25$ 个病人,$\bar{x}_2=79.3$ 次/分钟;医院 3,$n_3=16$ 个病人,$\bar{x}_3=80.1$ 次/分钟. 请计算这三个样本的总平均数.

解 ☞

$$总平均数=\frac{\sum_{i=1}^{3}n_i\bar{x}_i}{\sum_{i=1}^{3}n_i}=\frac{[(30\times76.2)+(25\times79.3)+(16\times80.1)]次/分钟}{30+25+16}$$

$$=\frac{5,550.1次/分钟}{71}=78.2次/分钟$$

6.11 几何平均

n 个正数 $x_1,x_2,\cdots,x_n$ 的**几何平均数**定义为这 n 个数的乘积的主 n 次方根(见 1.7 节),即

$$几何平均数=\sqrt[n]{x_1,x_2\cdots x_n}=\sqrt[n]{\prod_{i=1}^{n}x_i} \tag{6.20}$$

例 6.8 计算下列样本的算术平均和几何平均:1,3,5,6,8.

解 ☞

$$\bar{x}=\frac{\sum_{i=1}^{5}x_i}{5}=\frac{23}{5}=4.6$$

$$几何平均数=\sqrt[5]{\prod_{i=1}^{5}x_i}=\sqrt[5]{1\times3\times5\times6\times8}=\sqrt[5]{720}=3.727919,或 3.7$$

6.12 调和平均

n 个数据 $x_1,x_2,\cdots,x_n$ 的**调和平均数**定义为这 n 个数的倒数的算术平均数的倒数,即

$$调和平均数=\frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}} \tag{6.21}$$

例 6.9　计算例 6.8 中样本的调和平均数.

解 ☞　$调和平均数=\frac{1}{\frac{1}{5}\left(\frac{1}{1}+\frac{1}{3}+\frac{1}{5}+\frac{1}{6}+\frac{1}{8}\right)}$

$=\frac{1}{0.2(1+0.333333+0.2+0.166667+0.125)}=2.73973$，或 2.7

6.13　中位数和其他分位数

一组数据 $x_1,x_2,\cdots,x_n$ 排成阵列(见 4.1 节)后，数据中将阵列分成两个相等部分的值就称为这组数据的**中位数**. 数据中小于中位数和大于中位数的数据个数相等. 在一个阵列中确定中位数可以使用如下的**奇偶准则**:

如果阵列中数据个数为**奇数**，那么中位数是阵列的**中值**；如果阵列中数据个数为**偶数**，那么中位数是阵列的**两个中值的算术平均**.

数据的频数直方图或频数曲线中，经过中位数这点的垂直线将直方图或曲线分成两个相等的部分(见图 6-1). 对于中位数，没有通用的记号，这里我们使用现在统计书中相当常见的记号:$\tilde{x}$ 表示样本中位数，$\tilde{\mu}$ 表示总体中位数.

中位数是**分位数**的一种. 分位数是根据数据的升序阵列计算得到的特征数. 用符号 $Q_{j/m}$ 表示的分位数是指处于升序阵列的第 j/m 位置的值，也就是说，数据中有 j/m 的测量值小于这个值. 例如，有 1/2 的数据小于分位数 $Q_{2/4}$，有 1/4 的数据低于分位数 $Q_{1/4}$. 在频数直方图或频数曲线中，如果过 $Q_{j/m}$ 作一条垂线，那么这条垂线的左边图形的面积为总面积的 j/m. 这章中，我们讨论三种分位数:**四分位数**、**十分位数**和**百分位数**.

四分位数有 3 个:**第 1 四分位数**(用 $Q_{1/4}$ 或 Q_1 表示)，**第 2 四分位数**(用 $Q_{2/4}$ 或 Q_2 表示)，**第 3 四分位数**(用 $Q_{3/4}$ 或 Q_3 表示). 这些四分位数一起将阵列、频数直方图和频数曲线分为 4 个相等的部分. 十分位数有 9 个:**第 1 十分位数**(用 $Q_{1/10}$ 或 D_1 表示)，**第 2 十分位数**(用 $Q_{2/10}$ 或 D_2 表示)，…，**第 9 十分位数**(用 $Q_{9/10}$ 或 D_9 表示). 这些十分位数一起将阵列、频数直方图和频数曲线分为 10 个相等的部分. 百分位数共有 99 个:**第 1 百分位数**(用 $Q_{1/100}$ 或 P_1 表示)，**第 2 百分位数**(用 $Q_{2/100}$ 或 P_2 表示)，…，**第 99 百分位数**(用 $Q_{99/100}$ 或 P_{99} 表示). 它们一起将阵列、频数直方图和频数曲线分为 100 个相等的部分.

中位数和其他分位数是数据的**相对位置**($Q_{j/m}$ 相对于阵列或分布的边界的位置)的度量. 因为中位数表示的是阵列和分布的精确中点或重心的位置，所以它是集中趋势或中心位置的一种度量. 由于数据在中心位置附近的聚集特点，所以中位数与算术平均数一样，也是平均值的一种度量. 但是算术平均数是由所有数据计算得到的，会受到极端数据的影响，而中位数只涉及到值的排序，不会受极端值的影响. 所以中位数常被用来度量偏斜数据的平均值. 次序水平、间隔水平和比例水平数据都可以计算中位数.

例 6.10　请问中位数等于哪些分位数?

解 ☞　$\tilde{x}(或\ \tilde{\mu})=Q_2=D_5=P_{50}$

6.14　阵列的分位数计算公式

阵列的中位数通常采用奇偶准则计算，但阵列的分位数在不同统计书中有不同的计算方法和公式. 我们采用的计算阵列分位数的一般公式为

$$Q_{j/m}=x_i \tag{6.22}$$

其中 x_i 是指阵列中处于第 i 个位置的数值，$i=\frac{[j\times n]}{m}+\frac{1}{2}$(样本)或 $i=\frac{[j\times N]}{m}+\frac{1}{2}$(总体).

例 6.11 分别用 6.13 节中的奇偶准则和上面的(6.22)式，计算下列样本的中位数：(a) 12，13，11，(b) 12，13，14，15.

解 使用奇偶准则有：

(a) $n=3$，是奇数，所以中位数是阵列的中值：$\tilde{x}=13$.

(b) $n=4$，是偶数，所以中位数是阵列的两个中值的算术平均：$\tilde{x}=(13+14)/2=13.5$.

中位数也就是 $Q_{1/2}$，所以用(6.22)式有

(a) $i=\left[\frac{1\times 3}{2}\right]+\frac{1}{2}=2$，$\tilde{x}=x_2$，所以中位数是阵列中的第 2 个数值：13.

(b) $i=\left[\frac{1\times 4}{2}\right]+\frac{1}{2}=2.5$，$\tilde{x}=x_{2.5}$，所以中位数是第 2 个数值和第 3 个数值的中间值：13.5.

6.15 未分组频数分布的分位数计算公式

使用 6.13 节的奇偶准则或者(6.22)式计算得到的分位数可能落在**结值**(相等的值)的位置上. 许多统计书中，将结值作为中位数或其他分位数，这里有一个明显的问题：这种情形的分位数可能违背分位数的定义，并且许多分位数可能取相同的数值.

为了使未分组数据避免这些问题，许多统计书上建议：当分位数在结值时，使用如下**未分组频数分布的分位数计算公式**：

$$Q_{j/m}=b+\left[\frac{\frac{(j\times n)}{m}-Cf}{f}\right](w) \tag{6.23}$$

其中 b 是**分位数类**(分位数所在的类)的蕴含范围的下界，n 是样本容量(或 N 是总体容量)，Cf 是所有小于分位数类的累积频数，f 是分位数类的频数，w 是分位数类的蕴含范围的长度.

例 6.12 分别使用奇偶准则(见 6.13 节)、(6.22)式和(6.23)式，计算下列重量(以 lb 为单位)样本的中位数：1.1，1.2，1.2，1.3，1.3，1.3，1.3，1.3，1.4，1.5.

解 由于 $n=10$，所以由奇偶准则有 $\tilde{x}=(1.3+1.3)/2=1.3$ lb.

由(6.22)式，$i=\left[\frac{1\times 10}{2}\right]+\frac{1}{2}=5.5$，$\tilde{x}=x_{5.5}$，所以中位数是第 5 个数值(1.3 lb)和第 6 个数值(1.3 lb)的中间值：1.3 lb.

为了使用(6.23)式，我们先将样本转换成频数分布，见表 6.5，表 6.6 是其"小于式"累积频数分布. 我们常对这些分布用式(6.23)计算 $\tilde{x}=Q_{j/m}=Q_{1/2}$，根据分位数的定义知，有 1/2 的数据，即 5 个数值，必须小于 $Q_{1/2}$. 那么，由表 6.6"小于式"累积频数分布，我们可以看出，1.3 lb 是**分位数类**，现在也就是**中位数类**(中位数所在的类). 所以，$\tilde{x}$ 是 1.3 lb 的蕴含范围(1.25 lb 到 1.35 lb)中的某个值. 使用(6.23)式的前提假设是**分位数类中的数值在该类的蕴含范围中是均匀分布的**. 由于

$Q_{j/m}=\tilde{x}=Q_{1/2}$

$b=$ 中位数类的下界 $=1.25$ lb

$n=$ 样本容量 $=10$

$Cf=$ 所有小于中位数类的累积频数 $=3$

$f=$ 中位数类的频数 $=5$

$w=$ 中位数类的宽度 $=1.35\text{ lb}-1.25\text{ lb}=0.10$ lb，

所以

$$\begin{aligned}\tilde{x}=Q_{1/2}&=1.25\text{ lb}+\left[\frac{\frac{(1\times 10)}{2}-3}{5}\right](0.10\text{ lb})\\&=1.25\text{ lb}+(0.4\times 0.10\text{ lb})=1.29\text{ lb}\end{aligned}$$

表 6.5

重量(lb) x_i	频数 f_i
1.1	1
1.2	2
1.3	5
1.4	1
1.5	1
$\sum$	10

表 6.6

重量(lb)	累积频数
小于 1.1	0
小于 1.2	1
小于 1.3	3
小于 1.4	8
小于 1.5	9
小于 1.6	10

注:上述过程可简要描述为:我们发现中位数处于中位数类的蕴含范围的 2/5 的位置上,所以我们将该蕴含范围的长度乘以 0.4,然后加上蕴含范围的下界,得到的结果就是中位数.

例 6.13 对于表 6.1 中的 50 个长度数据(以 cm 为单位),用(6.23)式计算 Q_1,Q_2 和 Q_3.

解 在表 6.7 中,已经将表 6.1 的频数分布转换成了"小于式"累积频数分布. 因为 $Q_1 = Q_{1/4}$,所以有 1/4 的数据,即 12.5 个数值,必须小于 Q_1. 由"小于式"累积频数分布可以看出,**包含 Q_1 的类是** 1.4 cm,所以

$$Q_{j/m} = b + \left[\frac{\frac{(j\times n)}{m} - Cf}{f}\right](w)$$

$$Q_1 = Q_{1/4} = 1.35\ \text{cm} + \left[\frac{\frac{(1\times 50)}{4} - 9}{10}\right](0.10\ \text{cm})$$

$$= 1.35\ \text{cm} + (0.35 \times 0.10\ \text{cm})$$

$$= 1.385\ \text{cm},\text{或}\ 1.38\ \text{cm}$$

表 6.7

长度(cm)	累积频数
小于 1.2	0
小于 1.3	2
小于 1.4	9
小于 1.5	19
小于 1.6	31
小于 1.7	41
小于 1.8	48
小于 1.9	50

因为 $Q_2 = \tilde{x} = Q_{2/4}$,所以有 1/2 的数据,即 25 个数值必须小于 Q_2. 由"小于式"累积频数分布可以看出,**包含 Q_2 的类**(即中位数类)是 1.5 cm,所以

$$Q_2 = \tilde{x} = Q_{1/2} = 1.45\ \text{cm} + \left[\frac{\frac{(1\times 50)}{2} - 19}{12}\right](0.10\ \text{cm})$$

$$= 1.45\ \text{cm} + (0.50 \times 0.10\ \text{cm})$$

$$= 1.50\ \text{cm}$$

这与例 6.3 中所述相同,即单峰对称分布的算术平均数(1.5 cm)等于中位数.

因为 $Q_3 = Q_{3/4}$,所以有 3/4 的数据,即 37.5 个数值,必须小于 Q_3. 由"小于式"累积频数分布可以看出,**包含 Q_3 的类是** 1.6 cm,所以

$$Q_3 = Q_{3/4} = 1.55\ \text{cm} + \left[\frac{\frac{(3\times 50)}{4} - 31}{10}\right](0.10\ \text{cm})$$

$$= 1.55\ \text{cm} + (0.65 \times 0.10\ \text{cm})$$

$$= 1.615\ \text{cm},\text{或}\ 1.62\ \text{cm}$$

6.16 分组频数分布的分位数计算公式

如果样本或总体的所有测量值都已知,那么用奇偶准则或分位数计算公式得到的中位数都称为**精确的(或真实的)中位数**. 而根据分组频数分布计算得到的中位数只是精确中位数的一个近似值,所以称为**近似中位数**. **分组频数分布的分位数的计算公式为**

$$Q_{j/m} \approx b_c + \left[\frac{\frac{(j\times n)}{m} - Cf_c}{f_c}\right](w_c) \tag{6.24}$$

其中 $Q_{j/m}$ 是分位数,b_c 是分位数组(分位数所在的组)的下界,n 是样本容量(若是总体,则是总

体容量 N),Cf_c 是小于分位数组的所有组的累积频数,f_c 是分位数组的频数,w_c 是分位数组的组距.

例 6.14 表 6.2 是马拉松比赛中前 30 名男运动员成绩的未分组频数分布,表 6.3 是其分组频数分布. 请根据(6.23)式计算精确的中位数,根据(6.24)式计算近似中位数.

解 ☞ 因为 $\tilde{\mu}=Q_2$,所以有 1/2 的数据,即 15 个数值,必须小于 $\tilde{\mu}$. 由表 6.2 我们可以看出,包含 Q_2 的类是 142 min,所以精确的中位数为

$$Q_{j/m}=b+\left[\frac{\frac{(j\times N)}{m}-Cf}{f}\right](w)$$

$$Q_2 \quad \tilde{\mu}=Q_{1/2}=141.5\ \text{min}+\left[\frac{\frac{(1\times 30)}{2}-14}{4}\right](1.0\ \text{min})$$

$$=141.5\ \text{min}+(0.25\times 1.0\ \text{min})$$

$$=141.75\ \text{min},\text{或 } 141.8\ \text{min}$$

为了使用(6.24)式计算近似中位数,必须假定分位数组中所有数据在组中是均匀分布的. 下面计算近似中位数,首先要确定中位数组. 从表 6.3 看出,小于组 140—142 的数值有 11 个,小于组 143—145 的数值有 18 个,所以中位数组应是 140—142. 由于

b_c=中位数组的下界=139.5 min

N=总体容量=30

Cf_c=小于中位数组的所有组的累积频数=11

f_c=中位数组的频数=7

w_c=中位数组的组距=3.0 min

所以

$$Q_2=\tilde{\mu}=Q_{1/2}\approx 139.5\ \text{min}+\left[\frac{\frac{(1\times 30)}{2}-11}{7}\right](3.0\ \text{min})$$

$$\approx 139.5\ \text{min}+(0.571429\times 3.0\ \text{min})$$

$$\approx 141.214287\ \text{min},\text{或 } 141.2\ \text{min}$$

注:对于这个数据集,中位数的精确值是 141.8 min,而算术平均数的精确值是 139.9 min,中位数的近似值是 141.2 min,而算术平均数的近似值是 139.6 min. 这些结果说明,对于负偏态的分布,中位数位于算术平均数的右边. 这与例 6.4 的叙述相同.

6.17 中列数、四分位数中点和三点均值

中列数(或极差中点)是指数据集中最大值与最小值的算术平均数,即

$$\text{中列数}=\frac{x_s+x_l}{2} \tag{6.25}$$

四分位数中点是指第 1 四分位数与第 3 四分位数的算术平均数,即

$$\text{四分位数中值}=\frac{Q_1+Q_3}{2} \tag{6.26}$$

三点均值是指中位数和四分位数中值的算术平均数,即

$$\text{三点均值}=\frac{Q_2+\frac{Q_1+Q_3}{2}}{2} \tag{6.27}$$

或

$$\text{三点均值}=\frac{Q_1+2Q_2+Q_3}{4} \tag{6.28}$$

例 6.15 考虑表 6.1 中的长度数据,例 6.13 计算了该数据的四分位数. 请计算其中列数、四分位数中点和三点均值.

解 ☞ 中列数$=\frac{1.2\ \text{cm}+1.8\ \text{cm}}{2}=1.50\ \text{cm}$

$$四分位数中值=\frac{1.385\ \text{cm}+1.615\ \text{cm}}{2}=1.50\ \text{cm}$$

$$三点均值=\frac{1.385\ \text{cm}+2(1.50\ \text{cm})+1.615\ \text{cm}}{2}=1.50\ \text{cm}$$

注：对于对称的单峰分布，$\bar{x}=\tilde{x}$=中列数=四分位数中点=三点均值.

6.18　众数

一个数据集的**众数**是指数据集中出现次数最多的测量值. 如果在一个阵列中，有两个**相邻数值**的频数相同，并且它们的频数大于其他所有值的频数，那么众数一般由这两个相邻值的算术平均计算得到. 如果一个阵列中有两个**不相邻的数值**，它们的频数相同，并且大于其他所有值的频数，那么这两个数值都是众数. 如果一个数据集的所有值都有相同的频数，那么这个数据集的众数不存在.

例 6.16　计算下列样本的众数：(a) 2,3,3,3,4,5,6，(b) 2,3,4,5,6,6,6，(c) 2,3,3,4,4,4,5,5,6，(d) 2,3,3,3,4,4,4,5，(e) 2,3,3,3,4,5,6,6,6,8，(f) 2,2,3,3,4,4,5,5,6,6.

解 ☞　(a) 众数=3

(b) 众数=6

(c) 众数=4

(d) 众数=(3+4)/2=3.5

(e) 众数=3，众数=6

(f) 众数不存在

6.19　分组频数分布的众数计算公式

分组频数分布计算近似众数的方法有两种：(1)确定**众数组**(频数最高的组)，用该组的组中值作为近似众数；(2)对于**等组距**的分组分布，可以用下面的**分组频数分布的众数计算公式**计算：

$$众数\approx b_c+\left(\frac{d_1}{d_1+d_2}\right)(w_c) \tag{6.29}$$

其中 b_c 是众数组的下界，d_1 是众数组频数与其前一组频数之差，d_2 是众数组频数与其后一组频数之差，w_c 是众数组的组距.

例 6.17　附表 A.2 中 64 名同学期末考试成绩的分组分布见表 4.22，图 6-2 为其升序阵列茎叶表示图(见习题 6.18). 请根据图 6-2 计算精确众数，然后根据表 4.22 计算近似众数.

解 ☞　从图 6-2 可以看出频数最高的成绩是 90 分，所以众数=90.

从表 4.22 可以看出，众数组为 90—94，其组中值为 92，所以根据方法(1)有，众数≈92.

若使用方法(2)，从频数分布可以确定 $b_c=89.5$，$d_1=17-9=8$，$d_2=17-5=12$，$w_c=5$，所以

$$众数\approx 89.5+\left(\frac{8}{8+12}\right)(5)=89.5+(0.4\times 5)=91.5$$

4	9	(1)
5	5799	(4)
6	44457899	(8)
7	124688999	(9)
8	00012334444566777888	(20)
9	0000000111123344456789	(22)
		(64)

图 6-2

习题解答

算术平均数

6.1 利用(6.6)式计算下列总体的算术平均数:5.47×10^{-4} cm, 6.831×10^{-8} cm, 2.1211×10^{-5} cm.

解☞ $\mu=\frac{\sum x_i}{N}=\frac{5.47\times10^{-4}\text{ cm}+6.831\times10^{-8}\text{ cm}+2.1211\times10^{-5}\text{ cm}}{3}$

进行除法运算之前,将分子中的数转换成小数表示(见习题 2.15).

$$\mu=\frac{0.000547\text{ cm}+0.00000006831\text{ cm}+0.000021211\text{ cm}}{3}$$
$$=\frac{0.00056827931\text{ cm}}{3}$$
$$=\frac{0.000568\text{ cm}}{3}$$
$$=0.00018933333\text{ cm}$$

舍入成 0.000189 cm 或 1.89×10^{-4} cm.

6.2 利用例 6.1(a)中的数据证明:$\sum_{i=1}^{n}(x_i-\bar{x})=0$,其中平均数取 6 位数字.

解☞

$$\sum_{i=1}^{n}(x_i-\bar{x})=(1\text{g}-3.42857\text{g})+(3\text{g}-3.42857\text{g})+(2\text{g}-3.42857\text{g})$$
$$+(7\text{g}-3.42857\text{g})+(5\text{g}-3.42857\text{g})+(4\text{g}-3.42857\text{g})+(2\text{g}-3.42857\text{g})$$
$$=-2.42857\text{g}-0.42857\text{g}-1.42857\text{g}+3.57143\text{g}+1.57143\text{g}+0.57143\text{g}-1.42857\text{g}$$
$$=0.00001\text{g}$$

计算中使用 6 位数字的算术平均数,得到偏差之和在第 4 位小数上为零. 如果计算中使用 12 位数字的算术平均数 3.42857142857 g,那么

$$\sum_{i=1}^{n}(x_i-\bar{x})=0.00000000001\text{ g}$$

6.3 一家汽车经销处计划让 15 名销售员在 8 个星期内销售 150 辆新汽车. 在前 6 个星期,平均每星期销售汽车 $\bar{x}=19.5$ 辆. 请问剩下的两个星期内必须销售多少辆汽车才能完成任务?

解☞ 如果用 x_i 表示每星期销售汽车的总量,那么前 6 个星期销售汽车的总量为

$$\sum_{i=1}^{6}x_i=n\bar{x}=6\times19.5=117$$

因此,剩下的两个星期内必须销售 150−117=33 辆汽车才能完成任务.

6.4 利用(6.10)式,计算表 5.4 中样本的算术平均数.

解☞ 算术平均数的计算过程和结果见表 6.8.

表 6.8

重量(g) x_i	频数 f_i	f_ix_i(g)
14	2	28
15	2	30
16	4	64
17	18	306
18	24	432
19	35	665
20	5	100
$\sum$	90	1,625g

$\bar{x}=\frac{\sum f_ix_i}{n}=\frac{1,625\text{g}}{90}=18.1\text{g}$

注:该数据的相对频数直方图(见图 5-16)表明分布是负偏斜的. 但是该数据的算术平均数也是它的重心,尽管这个分布不如图 5-3 中的理想对称分布明显. 然而这并不表明图 5-16 中算术平均数的两边图形面积相等. 中位数将分布分成两个相等的部分,当我们计算该数据的中位数时(见习题 6.20),将会发现中位数位于算术平均数的右边,比算术平均数离偏斜的尾部更远.

6.5 利用(6.10)式,计算表 5.5 中样本的算术平均数.

解☞ 算术平均数的计算过程和结果见表 6.9.

注:对于这个正偏斜分布(见图 5-17),算术平均数仍然是

重心,但是现在它位于中位数的右边(见习题 6.21). 一般来说,对于偏斜分布,中位数总是比算术平均数更远离偏斜的尾部.

表 6.9

温度(℉) x_i	频数 f_i	f_ix_i(℉)
100	10	1,000
101	45	4,545
102	25	2,550
103	10	1,030
104	5	520
105	0	0
106	3	318
107	2	214
$\sum$	100	10,177℉

$$\bar{x}=\frac{\sum f_ix_i}{n}=\frac{10,177℉}{100}=101.8℉$$

由分组频数分布计算近似算术平均数

6.6 根据附表 A.2 中 64 个学生的期末考试成绩(第 3 列)直接计算它们的精确算术平均数. 然后再根据表 4.22 中的分组频数分布,利用(6.13)式计算它们的近似算术平均数.

解 ☞ 精确算术平均数为

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{5,221}{64}=81.6$$

利用(6.13)式计算近似算术平均数的过程和结果见表 6.10.

表 6.10

期末考试	组中值 m_i	频数 f_i	f_im_i
45—49	47	1	47
50—54	52	0	0
55—59	57	4	228
60—64	62	3	186
65—69	67	5	335
70—74	72	3	216
75—79	77	6	462
80—84	82	11	902
85—89	87	9	783
90—94	92	17	1,564
95—99	97	5	485
$\sum$		64	5,208

$$\bar{x}\approx\frac{\sum f_im_i}{n}=\frac{5,208}{64}=81.4$$

注:在计算算术平均数时,我们将这些离散比例水平数据看成连续比例水平测量值,精密度水平为计量单位. 于是结果舍入到十分位(见 6.3 节). 近似算术平均数(81.4)再次低估了精确算术平均数(81.6). 毫不奇怪,这个负偏态的分布(见图 5.25)的精确中位数及近似中位数(见习题 6.22)位于精确算术平均数及近似算术平均数的右边. 这对于中位数的图解法估计(84.5,见习题 5.29)也是正确的.

6.7 对于习题 4.17 中高尔夫球选手所获奖金,首先将奖金舍入到 \$1,000,计算奖金的精确算术平均数. 然后对于表 4.26 中的分组频数分布计算奖金的近似算术平均数. (在构造分组分布时,我们将奖金舍入到 \$1,000.)

解 ☞ 表 6.11 为数据舍入到 \$1,000 之后的未分组频数分布(为了节省空间,没有给出频数为零的类),利用(6.11)式计算总体的精确算术平均数 μ 的计算过程和结果也在表中给出了. 由于数据精确到千位,所以 μ 精确到百位. 利用(6.12)式计算总体的近似算术平均数的过程和结果见表 6.12.

注:对于不等组距的分组分布,可以使用与等组距分组分布相同的方法计算近似算术平均数. 注意近似算术平均数并不总是低估精确算术平均数,这里它高估了精确值. 由于计算近似算术平均数时需要所有组的组中值,所以**具有开端点组的分组频数分布无法计算近似算术平均数**. 在习题 6.23 将看到,该数据的中位数(精确值或近似值)位于算术平均数(精确值或近似值)的左边.

表 6.11

奖金($) x_i	频数 f_i	$f_i x_i$($)
2,000	18	36,000
3,000	7	21,000
4,000	5	20,000
6,000	7	42,000
8,000	7	56,000
10,000	3	30,000
15,000	7	105,000
22,000	5	110,000
30,000	3	90,000
36,000	2	72,000
40,000	1	40,000
45,000	1	45,000
83,000	3	249,000
200,000	1	200,000
$\sum$	70	$1,116,000

$$\mu=\frac{\sum f_i x_i}{N}=\frac{\$1,116,000}{70}=\$15,900$$

表 6.12

奖金($) x_i	组中值 m_i($)	频数 f_i	$f_i x_i$($)
2,000 4,000	3,000	30	90,000
5,000 10,000	7,500	17	127,500
11,000—20,000	15,500	7	108,500
21,000—30,000	25,500	8	204,000
31,000—40,000	35,000	3	106,500
41,000 45,000	43,000	1	43,000
46,000—81,000	63,500	0	0
82,000—83,000	82,500	3	247,500
84,000 198,000	141,000	0	0
199,000—200,000	199,500	1	199,500
$\sum$		70	$1,126,500

$$\mu\approx\frac{\sum f_i m_i}{N}=\frac{\$1,126,500}{70}=\$16,100$$

由编码数据计算算术平均数

6.8 首先直接计算下列重量测量值(以 g 为单位)样本的算术平均数:22,000.0,30,000.0,29,000.0,27,500.0,25,500.0,24,000.0. 然后根据式(6.14)、(6.15)和(6.16),并设 $a=0$ g,$b=\frac{1}{10,000}$,计算这些数据的算术平均数.

解 $\bar{x}$ 的直接计算和使用编码、解码公式的计算过程和结果见表 6.13.

表 6.13

重量(g) x_i	$c_i=0.0001x_i$ g
22,000.0	2.20000
24,000.0	2.40000
25,500.0	2.55000
27,500.0	2.75000
29,000.0	2.90000
30,000.0	3.00000
$\sum$ 158,000.0 g	15.80000 g

$$\bar{x}=\frac{\sum x_i}{n}=\frac{158,000.0\text{ g}}{6}=26,333.33\text{ g}$$

$$\bar{c}=\frac{\sum c_i}{n}=\frac{15.80000\text{ g}}{6}=2.633333\text{ g}$$

$$\bar{x}=\frac{1}{b}(\bar{c}-a)=\frac{1}{0.0001}(2.633333\text{ g})=26,333.33\text{ g}$$

6.9 两家衣服制造工厂的职工的平均小时工资为 $\bar{x}=\$5.39$. 两家工厂今年的效益一样好，他们决定增加职工的工资. 工厂 A 将给每位职工增加 5%的工资，而工厂 B 将给每位职工的小时工资增加 \$0.05. 请利用(6.14)式计算两家工厂职工的新的平均小时工资. 请问哪家工厂更慷慨大方?

解☞ 用 x_i 表示以前的小时工资. 对于工厂 A，现在每个工人的小时工资为 $c_i=\$0.00+1.05x_i$，因此，新的平均数为

$$\bar{c}=1.05\bar{x}=1.05(\$5.39)=\$5.66$$

对于工厂 B，现在每个工人的小时工资为 $c=\$0.05+x_i$，因此，新的平均数为

$$\bar{c}=\$0.05+\bar{x}=\$0.05+\$5.39=\$5.44$$

显然，工厂 A 更慷慨大方.

其他平均数:加权平均数、总平均数、几何平均数和调和平均数

6.10 生物课的最后成绩采用 100 分制，它由三部分构成:实验成绩占 25%，两次平时考试成绩占 25%，最后的考试成绩占 50%. 实验成绩采用 100 分制，平时考试成绩采用 50 分制，最后考试成绩采用 100 分制. 有一个学生实验成绩为 75 分，两次平时考试成绩分别为 40 分和 38 分，最后考试成绩为 85 分. 请用(6.17)式计算该学生生物课的最后成绩.

解☞ 已知 $k=4$，x_i 为各部分成绩，w_i 为各部分成绩所占的比重，因此

$$\bar{x}=\frac{\sum_{i=1}^{4}w_i\bar{x}_i}{\sum_{i=1}^{4}w_i}=\frac{0.25(75)+0.25(40+38)+0.50(85)}{0.25+0.25+0.50}=\frac{80.75}{1.0}=80.75$$

舍入成 80.8.

6.11 研究一种新型杂交水稻，要了解它的成熟日期(从种植到第一根成熟稻穗可以采摘的时间). 为此在四块不同的田地种植该种水稻，从每块田地随机选择 100 棵水稻测量它们的成熟日期. 计算每个样本的算术平均数，得到:$\bar{x}_1=70.1$ 天，$\bar{x}_2=71.3$ 天，$\bar{x}_3=69.5$ 天，$\bar{x}_4=69.2$ 天. 请计算这四个平均数的总平均数.

解☞ 根据(6.19)式和求和符号的性质(见习题 1.42 和 1.43)有

$$\text{总平均数}=\frac{\sum_{i=1}^{4}n_i\bar{x}_i}{\sum_{i=1}^{4}n_i}=\frac{100\sum_{i=1}^{4}\bar{x}_i}{400}=\frac{28,010}{400}=70.0\text{ 天}$$

6.12 计算下列样本的算术平均数和几何平均数:(a) 1,1,1,2,3,8,14,(b) 2,2,2,2,2,(c) 1,3,5,9,9,9,9.

解☞ (a) $\bar{x}=\frac{\sum_{i=1}^{7}x_i}{7}=\frac{30}{7}=4.3$

$$\text{几何平均数}=\sqrt[7]{\prod_{i=1}^{7}x_i}=\sqrt[7]{1\times1\times1\times2\times3\times8\times14}$$
$$=\sqrt[7]{627}=627^{1/7}=2.534603\text{，或 }2.5$$

(b) $\bar{x}=\frac{\sum_{i=1}^{5}x_i}{5}=\frac{10}{5}=2.0$

$$\text{几何平均数}=\sqrt[5]{\prod_{i=1}^{5}x_i}=\sqrt[5]{2\times2\times2\times2\times2}$$
$$=\sqrt[5]{32}=32^{0.2}=2.0$$

(c) $\bar{x}=\frac{\sum_{i=1}^{7}x_i}{7}=\frac{45}{7}=6.4$

$$几何平均数=\sqrt[7]{\prod_{i=1}^{7}x_i}=\sqrt[7]{1\times3\times5\times9\times9\times9\times9}$$

$$=\sqrt[7]{98,415}=98,415^{0.142857}=5.167658,或\ 5.2$$

注：这个习题证明了有关几何平均数的几个结论：(1)若数据是正偏斜的，例如(a)，那么几何平均数(2.5)受数据偏斜的影响比算术平均数(4.3)更小；(2)若所有数据的取值相同，例如(b)，那么它们的算术平均数等于几何平均数；(3)若数据的取值不同，那么它们的算术平均数大于几何平均数.

6.13 请证明一个数据集的几何平均数是这些数据的常用对数的算术平均数的反对数.(有关对数的运算见 1.10 节.)然后用这种方法重新计算例 6.8 中数据的几何平均数.

解 对于数据集 $x_1,x_2,\cdots,x_n$，它们的常用对数为 $\log_{10}x_1$，$\log_{10}x_2$，…，$\log_{10}x_n$，这些对数的算术平均数 $\bar{x}_{\log}$ 为

$$\bar{x}_{\log}=\frac{1}{n}(\log_{10}x_1+\log_{10}x_2+\cdots+\log_{10}x_n)$$

用 G 表示这些数据的几何平均数，则

$$G=\sqrt[n]{\prod_{i=1}^{n}x_i}=\sqrt[n]{x_1x_2\cdots x_n}=(x_1x_2\cdots x_n)^{1/n}$$

两边取常用对数，得

$$\log_{10}G=\log_{10}[(x_1x_2\cdots x_n)^{1/n}]$$
$$=\frac{1}{n}(\log_{10}x_1+\log_{10}x_2+\cdots+\log_{10}x_n)$$
$$=\bar{x}_{\log}$$

所以几何平均数 $G=\bar{x}_{\log}$ 的反对数.

例 6.8 中数据的常用对数为：$\log_{10}1=0.0000$，$\log_{10}3=0.4771$，$\log_{10}5=0.6990$，$\log_{10}6=0.7782$，$\log_{10}8=0.9031$. 因此

$$\bar{x}_{\log}=\frac{0.0000+0.4771+0.6990+0.7782+0.9031}{5}=0.57148$$

几何平均数 $G=\bar{x}_{\log}$ 的反对数=3.728039，或 3.7，与例 6.8 结果一样.

6.14 测量四个圆柱体的直径 D 和高度 L(以 cm 为单位). 这些测量值是比例水平，因此对每一个圆柱体都可以计算两个比值：D/L 和 L/D. 这些圆柱体的 D/L 值为：2/10，5/10，2/10，5/10；L/D 值为：10/2，10/5，10/2，10/5. 请利用这两组比值证明：对于比值数据，几何平均数比算术平均数更合适.

解 令 $x_i=D/L$，$y_i=L/D$，那么这两组比值的算术平均数分别为

$$\bar{x}=\frac{\sum_{i=1}^{4}x_i}{4}=\frac{0.2+0.5+0.2+0.5}{4}=0.35$$

$$\bar{y}=\frac{\sum_{i=1}^{4}y_i}{4}=\frac{5+2+5+2}{4}=3.5$$

这两组比值的几何平均数分别为

$$x_i\ 的几何平均数=\sqrt[4]{\prod_{i=1}^{4}x_i}=\sqrt[4]{(0.2)(0.5)(0.2)(0.5)}=0.316228$$

$$y_i\ 的几何平均数=\sqrt[4]{\prod_{i=1}^{4}y_i}=\sqrt[4]{5\times2\times5\times2}=3.16228$$

D/L 是 L/D 的倒数，因此 D/L 的平均数也应该是 L/D 的平均数的倒数. 但对算术平均数并不成立，$0.35\neq\frac{1}{3.5}=0.285714$，即 $\bar{x}$ 不是 $\bar{y}$ 的倒数. 但是，对于几何平均数，

$$0.316228=\frac{1}{3.16228}$$

即 x 的几何平均数是 y 的几何平均数的倒数. 所以说对于比值数据，几何平均数比算术平均数更合适.

6.15 利用(6.21)式计算习题6.12中三个样本的调和平均数.

解 用H表示数据的调和平均数.

(a)
$$H=\frac{1}{\frac{1}{7}\left(\frac{1}{1}+\frac{1}{1}+\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{8}+\frac{1}{14}\right)}$$
$$=\frac{1}{0.142857(1+1+1+0.5+0.333333+0.125+0.071429)}$$
$$=1.73708 \quad 或 \quad H=1.7$$

(b)
$$H=\frac{1}{\frac{1}{5}\left(\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}\right)}$$
$$=\frac{1}{0.2(2.5)}$$
$$=2.0$$

(c)
$$H=\frac{1}{\frac{1}{7}\left(\frac{1}{1}+\frac{1}{3}+\frac{1}{5}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}\right)}$$
$$=\frac{1}{0.142857[1+0.333333+0.2+4(0.111111)]}$$
$$=3.53933 \quad 或 \quad H=3.5$$

注:将这些结果与习题6.12的结果相比较,可以得到如下结论:

如果所有数据的取值相同,那么调和平均数=几何平均数=算术平均数.

如果数据的取值不同,那么调和平均数<几何平均数<算术平均数.

中位数和其他分位数

6.16 分别使用6.13节中的奇偶准则和(6.22)式计算下列样本的中位数:12,12,12,14,17,19,21.

解 由于$n=7$,所以由奇偶准则有,$\tilde{x}=14$.

由(6.22)式,$i=\left(\frac{1\times7}{2}\right)+\frac{1}{2}=4$,所以中位数是阵列中的第4个数值:14.

6.17 利用(6.22)式计算下列样本的D_3和P_{67}:1,6,7,8,9,11,13,15,20,21,28.

解 $D_3=Q_{3/10}$,于是$i=\left(\frac{3\times11}{10}\right)+\frac{1}{2}=3.8$,所以$D_3$是阵列中第3个数值和第4个数值之间8/10位置的值:$7+0.8(8-7)=7.8$.

$P_{67}=Q_{67/100}$,于是$i=\left(\frac{67\times11}{100}\right)+\frac{1}{2}=7.87$,所以$P_{67}$是阵列中第7个数值和第8个数值之间87/100位置的值:$13+0.87(15-13)=14.74$,或14.7.

6.18 将下面的重量测量值(以mg为单位)样本用简单茎叶表示法表示,开始部分和叶使用1位数字,茎距为1 mg(见习题5.43):8.1,4.9,6.5,6.3,5.8,3.7,2.2,1.1,5.7,7.4,2.8,3.3,6.9,3.9,3.1,2.0,5.3,7.0,1.3,1.9,7.9,6.2,5.0,5.2,2.5,2.1,4.2,3.6,7.6,4.5. 然后将这个茎叶表示转换成一升序阵列,并利用(6.22)式计算D_4和Q_2.

解 图6-3(a)为所求的简单茎叶表示图,将每条茎上的叶按从小到大的顺序排列后得到升序阵列,见图6-3(b).

$D_4=Q_{4/10}$,于是$i=\left(\frac{4\times30}{10}\right)+\frac{1}{2}=12.5$,所以$D_4=x_{12.5}$是阵列中第12个数值和第13个数值中间位置的值. 从图6-3(b)得到$x_{12}=3.7$ mg,$x_{13}=3.9$ mg,故

$$D_4=\frac{3.7\text{mg}+3.9\text{mg}}{2}=3.80\text{mg}$$

$Q_2=\tilde{x}=Q_{1/2}$,于是$i=\left(\frac{1\times30}{2}\right)+\frac{1}{2}=15.5$,所以$Q_2$是阵列中第15个数值和第16个数值中间位置的值. 从图6-3(b)得到$x_{15}=4.5$ mg,$x_{16}=4.9$ mg,故

$$Q_2=\frac{4.5\text{mg}+4.9\text{mg}}{2}=4.70\text{mg}$$

(a)			(b)		
1	139	(3)	1	139	(3)
2	28051	(5)	2	01258	(5)
3	73916	(5)	3	13679	(5)
4	925	(3)	4	259	(3)
5	87302	(5)	5	02378	(5)
5	5392	(4)	6	2359	(4)
7	4096	(4)	7	0469	(4)
8	1	(1)	8	1	(1)
		(30)			(30)

图 6-3

6.19 将下面的长度测量值(以 mm 为单位)样本用简单茎叶表示法表示，开始部分使用 1 位数字，叶使用 2 位数字，茎距为 0.1 mm(见习题 5.44)：0.948，0.513，0.687，0.231，0.299，0.717，0.379，0.310，0.785，0.542，0.222，0.593，0.827，0.309，0.784，0.502，0.272，0.492，0.256，0.651，0.329，0.358，0.447，0.699，0.589. 然后将这个茎叶表示转换成一升序阵列，并利用(6.22)式计算 D_7 和 P_{13}.

解 图 6-4 的左侧图形为所要求的简单茎叶表示图，将每条茎上的叶按从小到大的顺序排列后得到升序阵列，见图 6-4 的右侧图形.

最初的茎叶表示图		变换后的阵列表示图		
2	31,99,22,72,56	2	22,31,56,72,99	(5)
3	79,10,09,29,58	3	09,10,29,58,79	(5)
4	92,47	4	47,92	(2)
5	13,42,93,02,89	5	02,13,42,89,93	(5)
6	87,51,99	6	51,87,99	(3)
7	17,85,84	7	17,84,85	(3)
8	27	8	27	(1)
9	48	9	48	(1)
				(25)

图 6-4

$D_7=Q_{7/10}$，于是 $i=\left(\frac{7\times25}{10}\right)+\frac{1}{2}=18$，所以 D_7 是阵列中第 18 个数值. 从图 6-4(b)得到 $D_7=x_{18}=0.651$ mm.

$P_{13}=Q_{13/100}$，于是 $i=\left(\frac{13\times25}{100}\right)+\frac{1}{2}=3.75$，所以 $P_{13}=x_{3.75}$ 是阵列中第 3 个数值和第 4 个数值之间 3/4 位置的值. 从图 6-3(b)得到 $x_3=0.256$ mm，$x_4=0.272$ mm，故 $P_{13}=0.256\text{ mm}+0.75(0.272\text{mm}-0.256\text{ mm})=0.2680$ mm.

表 6.14

重量(g)	累积频数
小于 14	0
小于 15	2
小于 16	4
小于 17	8
小于 18	26
小于 19	50
小于 20	85
小于 21	90

6.20 利用(6.23)式计算表 6.8 中重量测量值的 Q_1、Q_2 和 Q_3.

解 为了使用(6.23)式，我们先将表 6.8 中的频数分布转换成“小于式”累积频数分布，见表 6.14.

90 个数据中有 1/4 的数，即 22.5 个数值，必须小于 Q_1，由表 6.14 看出，Q_1 所在的类为 17 g. 因此

$$Q_{j/m}=b+\left[\frac{\frac{(j\times n)}{m}-Cf}{f}\right](w)$$

$$Q_1=Q_{1/4}=16.5\text{g}+\left[\frac{\frac{(1\times 90)}{4}-8}{18}\right](1.0\text{g})$$

$$=16.5\text{g}+(0.805556\times 1.0\text{g})$$

$$=17.305556\text{g}\quad 或\quad Q_1=17.3\text{g}$$

90 个数据中有 1/2 的数,即 45 个数值,必须小于 Q_2,由表 6.14 看出,Q_2 所在的类为 18 g. 因此

$$Q_2=\tilde{x}=Q_{1/2}=17.5\text{g}+\left[\frac{\frac{(1\times 90)}{2}-26}{24}\right](1.0\text{g})$$

$$=18.291667\text{g}\quad 或\quad Q_2=18.3\text{g}$$

这证实了习题 6.4 中的结论:负偏斜分布的中位数位于算术平均数(18.1 g)的右边.

90 个数据中有 3/4 的数,即 67.5 个数值,必须小于 Q_3,由表 6.14 看出,Q_3 所在的类为 19 g. 因此

$$Q_3=Q_{3/4}=18.5\text{g}+\left[\frac{\frac{(3\times 90)}{4}-50}{35}\right](1.0\text{g})$$

$$=18.5\text{g}+(0.5\times 1.0\text{g})$$

$$=19.0\text{g}$$

6.21 利用(6.23)式计算表 6.9 中温度测量值(℉)的 Q_2 和 P_{87}.

解 ☞ 先将表 6.9 中的频数分布转换成"小于式"累积频数分布,见表 6.15.

表 6.15

温度(℉)	累积频数
小于 100	0
小于 101	10
小于 102	55
小于 103	80
小于 104	90
小于 105	95
小于 106	95
小于 107	98
小于 108	100

100 个数据中有 1/2 的数,即 50 个数值,必须小于 Q_2,由表 6.15 看出,Q_2 所在的类为 101℉. 因此

$$Q_{j/m}=b+\left[\frac{\frac{(j\times n)}{m}-Cf}{f}\right](w)$$

$$Q_2=\tilde{x}=Q_{1/2}=100.5℉+\left[\frac{\frac{(1\times 100)}{2}-10}{45}\right](1.0℉)$$

$$=100.5℉+(0.888889\times 1.0℉)$$

$$=101.388889℉\quad 或\quad Q_2=101.4℉$$

这证实了习题 6.5 中的结论:正偏斜分布的算术平均数(101.8℉)位于中位数的右边.

100 个数据中有 87/100 的数,即 87 个数值,必须小于 P_{87},由表 6.15 看出,P_{87} 所在的类为 103℉. 因此

$$P_{87}=Q_{87/100}=102.5℉+\left[\frac{\frac{(87\times 100)}{100}-80}{10}\right](1.0℉)$$

$$=102.5℉+(0.7\times 1.0℉)$$

$$=103.2℉$$

由分组频数分布计算分位数

6.22 将附表 A.2 中 64 名学生的期末考试成绩(第 3 列)用简单茎叶表示,并排列成升序阵列,见图 6-5. 将该图看成未分组频数分布,利用(6.23)式计算 Q_2 的精确值. 然后再根据表 6.10 中的分组频数分布利用(6.24)式计算 Q_2 的近似值.

解 ☞ 64 个数据中有 1/2 的数,即 32 个数值,必须小于 Q_2,由图 6-5 可知,Q_2 所在的类为 84. 因此精确中位数为

$$Q_{j/m}=b+\left[\frac{\frac{(j\times n)}{m}-Cf}{f}\right](w)$$

4	9	(1)
5	5799	(4)
6	44457899	(8)
7	124688999	(9)
8	00012334444566777888	(20)
9	0000000111123344456789	(22)
		(64)

图 6 - 5

表 6.16

期末考试	累积频数
小于 45	0
小于 50	1
小于 55	1
小于 60	5
小于 65	8
小于 70	13
小于 75	16
小于 80	22
小于 85	33
小于 90	42
小于 95	59
小于 100	100

$$Q_2 = \tilde{x} = Q_{1/2} = 83.5 + \left[\frac{\frac{(1\times 64)}{2} - 29}{4}\right](1.0)$$

$$= 83.5 + (0.75 \times 1.0)$$

$$= 84.25 \quad 或\ Q_2 = 84.2$$

为了使用分组频数分布的分位数计算公式,我们先将表 6.10 中的分组频数分布转换成分组"小于式"累积频数分布,见表 6.16. 要计算近似中位数,首先要确定中位数组. 从表 6.16 看出,中位数组应是 80—84. 因此近似中位数为

$$Q_{j/m} \approx b_c + \left[\frac{\frac{(j\times n)}{m} - Cf_c}{f_c}\right](w_c)$$

$$Q_2 = \tilde{x} = Q_{1/2} \approx 79.5 + \left[\frac{\frac{(1\times 64)}{2} - 22}{11}\right](5.0)$$

$$\approx 79.5 + (0.909091 \times 5.0)$$

$$\approx 84.045455 \quad 或 \quad Q_2 \approx 84.0$$

注:在习题 6.6 中我们指出,对于这个负偏斜分布,根据图解法估计的中位数(84.5)比算术平均数(精确值:81.6;近似值:81.4)大. 从这个习题我们看到,中位数的精确值(84.2)和近似值(84.0)也是位于算术平均数的右边.

6.23 根据表 6.11 中高尔夫球选手所获奖金的未分组频数分布,利用(6.23)式计算 Q_1 和 Q_2 的精确值. 然后再根据表 6.12 中的分组频数分布利用(6.24)式计算 Q_1 和 Q_2 的近似值.

解 70 个数据中有 1/4 的数,即 17.5 个数值,必须小于 Q_1,由表 6.11 看出,Q_1 所在的类为 \$2,000. 因此 Q_1 的精确值为

$$Q_{j/m} = b + \left[\frac{\frac{(j\times n)}{m} - Cf}{f}\right](w)$$

$$Q_1 = Q_{1/4} = \$1,500 + \left[\frac{\frac{(1\times 70)}{4} - 0}{18}\right](\$1,000)$$

$$= \$1,500 + (0.972222 \times \$1,000)$$

$$= \$2,472.222 \quad 或\ Q_1 = \$2,500$$

70 个数据中有 1/2 的数,即 35 个数值,必须小于 Q_2,由表 6.11 看出,Q_2 所在的类为 \$6,000. 因此 Q_2 的精确值为

$$Q_2 = \tilde{\mu} = Q_{1/4} = \$5,500 + \left[\frac{\frac{(1\times 70)}{2} - 30}{7}\right](\$1,000)$$

$$= \$5,500 + (0.714286 \times \$1,000)$$

$$= \$6,214.286 \quad 或\ Q_2 = \$6,200$$

虽然表 6.12 中的分组频数分布具有不等组距,但是计算近似分位数的方法没有改变. 将分组频数分布转换成分组"小于式"累积频数分布,见表 6.17. 从表 6.17 看出,为求 Q_1 的近似值,有 17.5 个数值必须小于 Q_1,于是 Q_1 所在的组是 \$2,000—\$4,000,因此 Q_1 的近似值为

$$Q_{j/m} \approx b_c + \left[\frac{\frac{(j\times n)}{m} - Cf_c}{f_c}\right](w_c)$$

$$Q_1 = Q_{1/4} \approx \$1,500 + \left[\frac{\frac{(1\times 70)}{4} - 0}{30}\right](\$3,000)$$

$$\approx \$1,500 + (0.583333 \times \$3,000)$$

$$\approx \$3,249.999 \quad 或 Q_1 \approx \$3,200$$

为求 Q_2 的近似值,有 35 个数值必须小于 Q_2,于是 Q_2 所在的组是 \$5,000—\$10,000,因此 Q_2 的近似值为

$$Q_2 = \tilde{\mu} = Q_{1/2} \approx \$4,500 + \left[\frac{\frac{(1\times 70)}{2} - 30}{17}\right](\$6,000)$$

$$\approx \$4,500 + (0.294118 \times \$6,000)$$

$$\approx \$6,264.708 \quad 或 Q_2 \approx \$6,300$$

表 6.17

奖金(\$)	累积频数
小于 1,500	0
小于 4,500	30
小于 10,500	47
小于 20,500	54
小于 30,500	62
小于 40,500	65
小于 45,500	66
小于 81,500	66
小于 83,500	69
小于 198,500	69
小于 200,500	70

注:这个正偏斜分布的中位数(精确值:\$6,200;近似值:\$6,300)位于习题 6.7 中的算术平均数(精确值:\$15,900;近似值:\$16,100)的左边.

6.24 根据高尔夫球选手所获奖金的"小于式"累积百分比曲图(见图 5-38),利用图解法估计 Q_1 和 Q_3.

解 将图 5-38 的相关部分放大,见图 6-6. 为了估计 Q_1,过 Y 轴的 25%的点作水平线,与累积

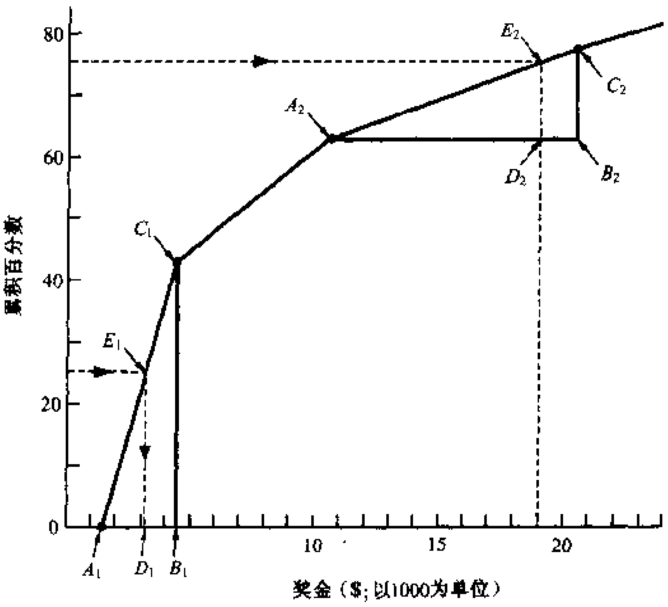


图 6-6

曲线相交于点 E_1. 然后过点 E_1 作垂直线,与 X 轴相交于点 D_1. 点 D_1 的横坐标大约是 \$3,000,可以看成 Q_1 的粗略估计. 更精确的估计可以通过相似三角形 $A_1B_1C_1$ 和 $A_1D_1E_1$ 计算出来. 这两个三角形的边长具有如下关系:

$$\frac{A_1D_1}{A_1B_1} = \frac{D_1E_1}{B_1C_1}$$

已知 $A_1B_1 = Q_1$ 所在组的组距 = \$3,000,$B_1C_1 = 42.8571 - 0.0 = 42.8571$,$D_1E_1 = 25.0 - 0.0 = 25.0$,因此可以解出 A_1D_1:

$$\frac{A_1D_1}{\$3,000}=\frac{25.0}{42.8571}$$

$$A_1D_1=\$3,000(0.583334)=\$1,750.00$$

因此

$$Q_1=\$1,500+\$1,750=\$3,250 \quad \text{或} \quad Q_1\approx\$3,200$$

为了估计 Q_3，过 Y 轴的 75%的点作水平线，与累积曲线相交于点 E_2. 然后过点 E_2 作垂直线，与 X 轴相交的点的横坐标大约是 \$19,000，可以看成 Q_3 的粗略估计. 更精确的估计可以通过相似三角形 $A_2B_2C_2$ 和 $A_2D_2E_2$ 计算出来. 这两个三角形的边长具有如下关系：

$$\frac{A_2D_2}{A_2B_2}=\frac{D_2E_2}{B_2C_2}$$

已知 $A_2B_2=Q_3$ 所在组的组距 $=\$10,000$，$B_2C_2=77.1429-62.1429=15.0$，$D_2E_2=75.0-62.1429=12.8571$，因此可以解出 A_2D_2：

$$\frac{A_2D_2}{\$10,000}=\frac{12.8571}{15.0}$$

$$A_2D_2=\$10,000(0.857140)=\$8,571.40$$

因此

$$Q_3=\$10,500+\$8,571.40=\$19,071.40 \quad \text{或} \ Q_3=\$19,000$$

中列数、四分位数中点和三点均值

6.25 根据表 6.8 中的重量测量值和习题 6.20 中计算得到的分位数，确定它们的中列数、四分位数中点和三点均值.

解 ☞ 中列数 $=\frac{x_s+x_l}{2}=\frac{14\text{g}+20\text{g}}{2}=17.0\text{g}$

四分位数中点 $=\frac{Q_1+Q_3}{2}$

$=\frac{17.305556\text{g}+19.0\text{g}}{2}$

$=18.152778\text{g}$，或 18.2g

三点均值 $=\frac{Q_1+2Q_2+Q_3}{4}$

$=\frac{17.305556\text{g}+2(18.291667)+19.0\text{g}}{4}$

$=18.222222\text{g}$ 或 18.2g

可以看出，对于负偏斜分布有，

$$\text{中列数}<\bar{x}<\text{四分位数中点}<\text{三点均值}<\tilde{x}$$

6.26 根据表 6.9 中的温度测量值和习题 6.21 和习题 6.51 中计算得到的分位数确定它们的中列数、四分位数中点和三点均值.

解 ☞ 中列数 $=\frac{x_s+x_l}{2}=\frac{100^\circ\text{F}+107^\circ\text{F}}{2}=103.5^\circ\text{F}$

四分位数中点 $=\frac{Q_1+Q_3}{2}$

$=\frac{100.833333^\circ\text{F}+102.3^\circ\text{F}}{2}$

$=101.566666^\circ\text{F}$，或 101.6°F

三点均值 $=\frac{Q_1+2Q_2+Q_3}{4}$

$=\frac{100.833333^\circ\text{F}+2(101.388889^\circ\text{F})+102.3^\circ\text{F}}{4}$

$=101.477778^\circ\text{F}$，或 101.5°F

可以看出，对于正偏斜分布有，

$$\tilde{x}<\text{三点均值}<\text{四分位数中点}<\bar{x}<\text{中列数}$$

6.27 根据表 6.11 中高尔夫球选手获得的奖金数据和习题 6.23 和习题 6.53 中计算得到的

分位数的精确值确定它们的中列数、四分位数中点和三点均值.

解 中列数$=\frac{x_s+x_l}{2}=\frac{\$2,000+\$200,000}{2}=\$101,000$

四分位数中点$=\frac{Q_1+Q_3}{2}$

$=\frac{\$2,472.222+\$15,285.714}{2}$

$=\$8,878.968$，或　$\$8,900$

三点均值$=\frac{Q_1+2Q_2+Q_3}{4}$

$=\frac{\$2,472.222+2(\$6,214.286)+\$15,285.714}{4}$

$=\$7,546.627$　或　$\$7,500$

可以看出，对于由正偏斜分布计算得到的精确值有，

$$\tilde{\mu}<\text{三点均值}<\text{四分位数中点}<\mu<\text{中列数}$$

6.28 根据表 6.12 中高尔夫球选手获得的奖金的分组分布和习题 6.23 和习题 6.53 中计算得到的分位数的近似值确定它们的中列数、四分位数中点和三点均值.

解 如果只有分组数据，例如表 6.12 中的分组频数分布，那么用来估计 x_s 和 x_l 的方法有两种：(1)用最小值所在组的组中值估计 x_s，用最大值所在组的组中值估计 x_l；(2)用最小值所在组的下界估计 x_s，用最大值所在组的上界估计 x_l；

因为中列数$=\frac{x_s+x_l}{2}$，所以用方法(1)有

中列数$\approx\frac{\$3,000+\$199,500}{2}=\$101,250$，或$\$101,200$

用方法(2)有

中列数$\approx\frac{\$1,500+\$200,500}{2}=\$101,000$

因为四分位数中点$=\frac{Q_1+Q_3}{2}$，根据习题 6.23 和习题 5.53 中计算得到的分位数的近似值有

四分位数中点$\approx\frac{\$3,249.999+\$18,357.140}{2}=\$10,803.5695$，或$\$10,800$

因为三点均值$=\frac{Q_1+2Q_2+Q_3}{4}$，根据习题 6.23 和习题 6.53 中计算得到的分位数的近似值有

三点均值$\approx\frac{\$3,249.999+2(\$6,264.708)+\$18,357.140}{4}=\$8,534.139$ 或 $\$8,500$

可以看出，对于由正偏斜分布计算得到的近似值有，

$$\tilde{\mu}<\text{三点均值}<\text{四分位数中点}<\mu<\text{中列数}$$

这与习题 6.27 中的排列顺序相同.

众数

6.29 对于下列未分组分布，首先计算它们的众数，然后将 $\bar{x}$、$\tilde{x}$ 和众数按大小顺序排列：(a) 表 6.1，(b) 表 6.8，(c) 表 6.9.

解 (a) 对于表 6.1 中的分布，众数＝1.5 cm. 从表 6.1 和例 6.13 可知，$\bar{x}=1.50$ cm，$\tilde{x}=1.50$ cm. 所以，对于这个单峰对称分布，$\bar{x}=\tilde{x}=$众数.

(b) 对于表 6.8 中的分布，众数＝19 g. 从表 6.8 和习题 6.20 可知，$\bar{x}=18.1$ g，$\tilde{x}=18.3$ g. 所以，对于这个单峰负偏斜分布，$\bar{x}<\tilde{x}<$众数.

(c) 对于表 6.9 中的分布，众数＝101℉. 从表 6.9 和习题 6.21 可知，$\bar{x}=101.8$℉，$\tilde{x}=101.4$℉. 所以，对于这个单峰正偏斜分布，众数$<\tilde{x}<\bar{x}$.

注：该习题表明，如果分布是单峰对称的，那么 $\bar{x}=\tilde{x}=$众数；如果分布是单峰偏斜的，那么这三个统计测量值是不同的，$\bar{x}$ 最靠近偏斜尾部的极端值，众数最远离偏斜尾部，$\tilde{x}$ 位于 $\bar{x}$ 和众数之间.

6.30 确定下列数据的 $\bar{x}$，$\tilde{x}$ 和众数，并进行排序：

$$x_1 = x_2 = x_3 < x_4 < x_5 < x_6 < x_7.$$

解 $\bar{x} = \dfrac{\sum_{i=1}^{7} x_i}{7}$，$\tilde{x} = x_4$，众数 $= x_1 = x_2 = x_3$

这是个正偏斜分布，由习题 6.29(c)可知，众数$<\tilde{x}<\bar{x}$.

6.31 许多统计书指出：对于"中等偏斜"的单峰频数分布，有下列关系：

$$\bar{x} - \tilde{x} = \frac{1}{3}(\bar{x} - \text{众数})$$

也就是说，$\bar{x}$ 和 $\tilde{x}$ 之间的距离是 $\bar{x}$ 到众数的距离的三分之一. 这个关系是通过许多数据集的验证得到的，而不是通过数学推导得到的，所以被称为**经验准则**. 计算下列样本的 $\bar{x}$、$\tilde{x}$ 和众数，然后检验这种关系是否正确：(a) 1.5，3.0，3.0，3.0，4.0，5.0，6.0，7.0，8.0，(b) 1.5，3.0，3.0，3.0，4.0，5.0，7.0，9.0，10.0.

解 (a) $\bar{x}=4.5$，由奇偶准则，$\tilde{x}=4.0$，众数$=3.0$

因此

$$\bar{x} - \tilde{x} = 4.5 - 4.0 = 0.5$$
$$\bar{x} - \text{众数} = 4.5 - 3.0 = 1.5$$
$$0.5 = \frac{1}{3}(1.5)$$

所以对于这个单峰的、适度正偏斜的分布，关系 $\bar{x}-\tilde{x}=\frac{1}{3}(\bar{x}-\text{众数})$成立.

(b) 众数和 $\tilde{x}$ 与(a)中的相同，但是这组数据的正偏斜程度增加了，因此 $\bar{x}$ 变得更大(向正方向移动)，计算得到 $\bar{x}=5.055556$，

因此

$$\bar{x} - \tilde{x} = 5.055556 - 4.0 = 1.055556$$
$$\bar{x} - \text{众数} = 5.055556 - 3.0 = 2.055556$$

所以对于这个正偏斜程度增加的分布，1/3 的关系不成立，它们的关系为

$$\bar{x} - \tilde{x} = \frac{1}{1.9473678}(\bar{x} - \text{众数})$$

注：通常，经验 1/3 准则对于单峰的、适度偏斜的频数分布是近似正确的. 但是，如果偏斜程度增加，$\bar{x}$ 向远离 $\tilde{x}$ 和众数的方向移动时，就会违反这个准则.

6.32 请问图 6-7 中的**双峰频数直方图**的众数是什么？

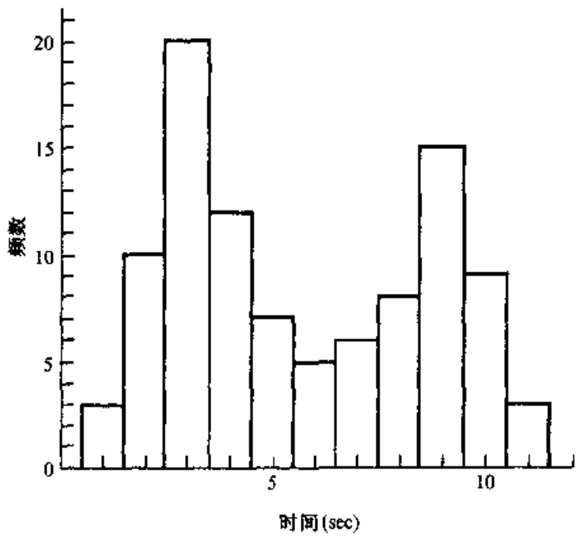


图 6-7

解 根据 6.18 节的定义"一个数据集的众数是指数据集中出现次数最多的测量值"，图 6-7 中的分布有一个众数：3 sec. 然而，根据例 5.2 的叙述，有两个峰的频数直方图称为双峰的. 这种看上

去的矛盾是如何解决的呢?

如果分布有两个峰,它们在直方图上有相同的高度(频数),那么分布有两个众数(见 6.18 节),这时称为双峰不存在矛盾. 但是,如果分布的两个不同的峰的高度不相同,如图 6-7,这时就存在矛盾. 按照惯例,这种分布也称为双峰分布,但许多统计书称这种分布只有一个**真实众数**. 而有的统计书则将较高峰的测量值称为**主要众数**,较低峰的测量值称为**次要众数**,或者称为**全局众数**和**局部众数**.

类似的,如果分布有三个不同峰的高度相同,那么分布有三个众数,这种分布称为**三峰分布**. 如果三个峰的高度不同,那么分布仍然称为三峰分布,但是我们称它有一个真实众数,或一个主要众数和两个次要众数,或一个全局众数和两个局部众数.

如习题 5.37 中指出的,双峰分布通常意味着样本来自两个不同的总体. 三峰分布则表明可能在取样或测量过程出现问题.

6.33　图 6-8 是四个分布的频数直方图,请根据图形从下列术语和关系中选择各分布的分布特征和统计量的关系:对称的,正偏斜的,负偏斜的,单峰的,双峰的,没有众数,$\sum(x_i-\bar{x})=0$,$\bar{x}=\tilde{x}$,$\bar{x}=\tilde{x}=$众数,众数$<\tilde{x}<\bar{x}$,$\bar{x}<\tilde{x}<$众数,$\bar{x}=$三点均值,$\bar{x}<$三点均值,三点均值$<\bar{x}$.

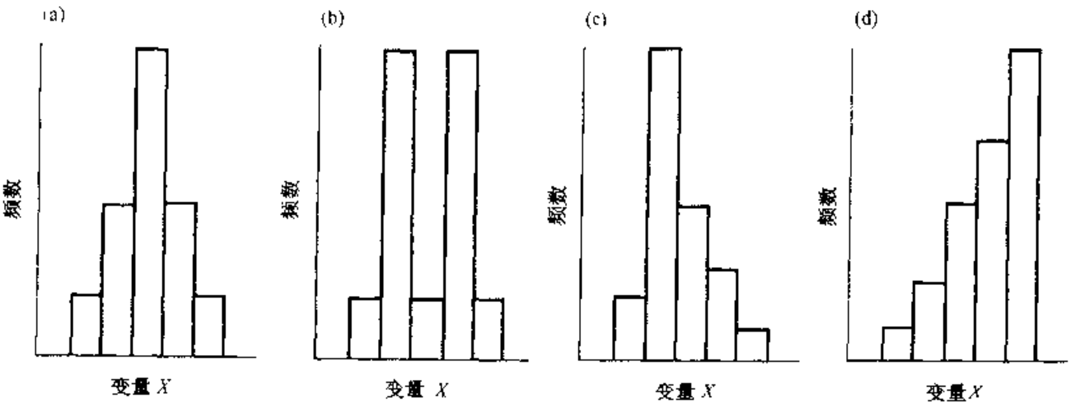


图 6-8

解　(a) 对称的,单峰的,$\sum(x_i-\bar{x})=0$,$\bar{x}=\tilde{x}=$众数,$\bar{x}=$三点均值.

(b) 对称的,双峰的,$\sum(x_i-\bar{x})=0$,$\bar{x}=\tilde{x}$,$\bar{x}=$三点均值.

(c) 正偏斜的,单峰的,$\sum(x_i-\bar{x})=0$,众数$<\tilde{x}<\bar{x}$,三点均值$<\bar{x}$.

(d) 负偏斜的,单峰的,$\sum(x_i-\bar{x})=0$,$\bar{x}<\tilde{x}<$众数,$\bar{x}<$三点均值.

6.34　图 6-9 是一组数据的频数直方图,请计算 $\bar{x}$、$\tilde{x}$ 和众数.

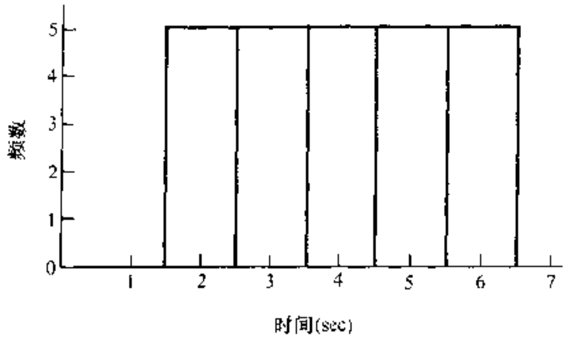


图 6-9

解 ☞

$$\bar{x}=\frac{\sum_{i=1}^{5}f_ix_i}{\sum_{i=1}^{5}f_i}$$

$$=\frac{(5\times 2\text{sec})+(5\times 3\text{sec})+(5\times 4\text{sec})+(5\times 5\text{sec})+(5\times 6\text{sec})}{5+5+5+5+5}$$

$$=4.0\text{sec}$$

根据(6.23)式

$$Q_{j,m}=b+\left[\frac{\frac{(j\times n)}{m}-Cf}{f}\right](w)$$

$$Q_2=\tilde{x}=Q_{1/2}=3.5\text{sec}+\left[\frac{\frac{(1\times 25)}{2}-10}{5}\right](1.0\text{sec})$$

$$=3.5\text{sec}+(0.5\times 1.0\text{sec})$$

$$=4.0\text{sec}$$

由于五个测量值的频数相同,所以这个分布没有众数.

6.35 下面给出几种本章讨论过的描述性度量的例子,请指出它们分别是哪种描述性度量:(a) 日平均气温,(b) 家庭平均收入,(c) 平均击球数.

解 ☞ (a) 日平均气温通常是当天的最低气温和最高气温的算术平均数,所以是中列数.

(b) 收入数据常常表现出正偏斜,由于算术平均数会受极端数据的影响而中位数却不会,所以家庭平均收入通常是中位数. 由于三点均值是从中位数及其他两个四分位数得到的,所以现在用三点均值作偏斜数据平均值的度量也越来越多.

(c) 平均击球数是算术平均数.

由分组频数分布计算众数

6.36 对于表 4.38 中的男生身高的分组频数分布,分别用 6.19 节的两种方法确定其众数的近似值.

解 ☞ 这个分布的众数组在表 4.38 中是(68.50 in 到 69.49 in),组中值是 69.00 in,因此,由方法(1)可知:众数≈69.00in.

根据方法(2)有

$$众数\approx b_i+\left(\frac{d_1}{d_1+d_2}\right)(w_i)$$

$$\approx 68.495\text{in}+\left(\frac{6}{6+5}\right)(1.000\text{in})$$

$$\approx 69.040455\text{in},或\ 69.040\text{in}$$

补 充 习 题

算术平均数

6.37 利用(6.3)式计算下列样本的算术平均数:(a) 0 sec,0 sec,0 sec,0 sec,0 sec,(b) 10 sec,0 sec,0 sec,0 sec,0 sec,(c) 10.127 km,11.963 km,112.217 km,9.777 km,13.833 km,14.542 km,(d) 10.127 km,11.963 km,0.007 km,9.777 km,13.833 km,14.542 km.

答案 (a) 0.0 sec,(b) 2.0 sec,(c) 28.7432 km,(d) 10.0415 km.

6.38 利用(6.6)式计算下列总体的算术平均数:100,000 只猪,115,100 只猪,152,643 只猪.

答案 进行舍入运算后为 123,000 只猪,或 1.23×10^5 只猪.

6.39 已知 $n=5$,$\bar{x}=14.6$,$\sum_{i=1}^{4}x_i=58$,请问 x_5 等于多少?

答案　$\sum_{i=1}^{5}x_i = 5\times 14.6 = 73$，因此 $x_5=73-58=15$.

6.40　利用(6.10)式,计算表 5.6 中样本的算术平均数.

答案　18.0

由分组频数分布计算近似算术平均数

6.41　利用(6.13)式计算的表 4.20 中长度的分组频数分布的近似算术平均数.

答案　$\bar{x}\approx 1.72$ mm

由编码数据计算算术平均数

6.42　首先直接计算下列重量测量值(以 g 为单位)样本的算术平均数:7.77×10^{-6},7.72×10^{-6},7.74×10^{-6},7.73×10^{-6},7.79×10^{-6},7.75×10^{-6}.(含有用科学计数法表示的数的计算过程见习题 2.21 和 2.22.)然后根据式(6.14)、(6.15)和(6.16),并设 $a=-770$ g,$b=10^8$,计算这些数据的算术平均数.

答案　$\bar{x}$ 的直接计算和使用编码、解码公式的计算过程和结果见表 6.18.

表 6.18

重量(g) x_i	$c_i=-770\text{g}+10^8x_i\text{g}$
7.72×10^{-6}	2
7.73×10^{-6}	3
7.74×10^{-6}	4
7.75×10^{-6}	5
7.77×10^{-6}	7
7.79×10^{-6}	9
$\sum$ 0.00004650g	30g

$$\bar{x}=\frac{\sum x_i}{n}=\frac{0.00004650\text{g}}{6}=0.000007750\text{g}=7.750\times10^{-6}\text{g}$$

$$\bar{c}=\frac{\sum c_i}{n}=\frac{30\text{g}}{6}=5.0\text{g}$$

$$\bar{x}=\frac{1}{b}(\bar{c}-a)=\frac{1}{10^8}[5.0\text{g}-(-770\text{g})]=\frac{775.0\text{g}}{10^8}=7.750\times10^{-6}\text{g}$$

其他平均数:加权平均数、总平均数、几何平均数和调和平均数

6.43　在 Colorado 大学,学生的总平均成绩(GPA)的计算公式为

$$\text{GPA}=\frac{\sum_{i=1}^{k}w_ix_i}{\sum_{i=1}^{k}w_i}$$

其中 w_i 表示学分,x_i 表示每个学分的成绩,k 是课程数.每个学分的成绩采用下面的计分系统:$A=4.0$,$A^-=3.7$,$B^+=3.3$,$B=3.0$,$B^-=2.7$,$C^+=2.3$,等等.一名新生选修了四门课程,其成绩分别为:课程 1:A,4 学分;课程 2:B^+,6 学分;课程 3:C^+,2 学分;课程 4:B^-,3 学分,请计算他的总平均成绩.

答案

$$\text{GPA}=\frac{\sum_{i=1}^{k}w_ix_i}{\sum_{i=1}^{k}w_i}$$

$$=\frac{(4\times4.0)+(6\times3.3)+(2\times2.3)+(3\times2.7)}{4+6+2+3}$$

$$=\frac{48.5}{15}=3.23$$

6.44 某试验的三次结果如下：$\bar{x}_1=19.2$ cm，$n_1=10$；$\bar{x}_2=17.4$ cm，$n_2=15$；$\bar{x}_3=18.5$ cm，$n_3=18$. 请计算这些结果的总平均数.

答案 18.2 cm

6.45 对于样本：9，9，11，7，计算 (a) 算术平均数，(b) 几何平均数，(c) 调和平均数.

答案 (a) 9.0，(b) 8.0，(c) 8.8.

中位数和其他分位数

6.46 分别使用6.13节中的奇偶准则和(6.22)式计算下列样本的中位数：12，12，12，12，13，13，13，13.

答案 由奇偶准则有 $\tilde{x}=12.5$；由(6.22)式，$\tilde{x}=12.5$

6.47 利用(6.22)式计算习题6.17中样本的 Q_1 和 $\tilde{x}$.

答案 $Q_1=7.25$ 或 7.2，$\tilde{x}=11$

6.48 根据(6.22)式和图6-3计算习题6.18中样本的 P_{82}.

答案 6.91 mg

6.49 根据(6.22)式和图6-4计算习题6.19中样本的 Q_3.

答案 0.690 mg

6.50 根据(6.23)式计算表6.8和表6.14中重量数据的 D_7.

答案 17.91667 g 或 17.9 g

6.51 根据(6.23)式计算表6.9和表6.15中温度数据的 Q_1 和 Q_3.

答案 $Q_1=100.833333$°F，或 100.8°F；$Q_3=102.3$°F

由分组频数分布计算分位数

6.52 根据图6-5中的考试成绩，用(6.23)式计算 P_{77} 的精确值，然后根据表6.10和表6.16中的分组频数分布，利用(6.24)式计算 P_{77} 的近似值.

答案 精确值为89.9，近似值为90.3.

6.53 根据表6.11中高尔夫球选手所获奖金的未分组频数分布，利用(6.23)式计算 Q_3 的精确值. 然后再根据表6.12和表6.17，利用(6.24)式计算 Q_3 的近似值.

答案 精确值为＄15,285.714 或＄15,300，近似值为＄18,357.140 或＄18,400

中列数、四分位数中点和三点均值

6.54 根据习题6.18给出的重量测量值和图6-3，可知：$x_1=1.1$ mg，$x_n=8.1$ mg，$Q_2=4.70$ mg. 首先确定 Q_1 和 Q_3，然后计算它们的中列数、四分位数中点和三点均值.

答案 $Q_1=2.8$ mg，$Q_3=6.3$ mg，中列数=4.60 mg，四分位数中点=4.55 mg，三点均值=4.62 mg

众数

6.55 分别确定表4.17的女生头发颜色样本和表4.18的男生头发颜色样本的 $\bar{x}$、$\tilde{x}$ 和众数.

答案 尽管对于名义水平数据不可能确定 $\bar{x}$ 和 $\tilde{x}$，但是可以确定众数. 对于女生，众数=金色；对于男生，众数=棕色. 这类众数也称为**众数类**.

6.56 确定下列数据的 $\bar{x}$、$\tilde{x}$ 和众数，并进行排序：

$$x_1<x_2<x_3<x_4=x_5.$$

答案 $\bar{x}=\dfrac{\sum_{i=1}^{5}x_i}{5}$，$\tilde{x}=x_3$，众数 $=x_4=x_5$. 由习题6.29(b)可知，对于单峰负偏斜分布有 $\bar{x}<\tilde{x}<$ 众数.

第七章　描述性统计：离散性度量

7.1　极差作为一种离散性度量为什么具有有限值

第六章研究了集中趋势的均值和位置描述性度量——用一个数值概括数据集的这些特征.本章我们将讨论数据集另外的一种特征描述性度量：数据是如何散布的，这与第六章介绍的度量有关.我们称这种度量为**离散性度量**，**变差的度量**，或**变异性的度量**.

为理解这种离散性度量的必要，考虑图 7-1 所示的两条频数曲线，它们都是单峰对称的，具有相同的均值和中位数，但是其中一条曲线在均值两侧陡峭地上升，而另一条曲线在均值处的密集程度较弱，更多地表现为离中向外趋势.

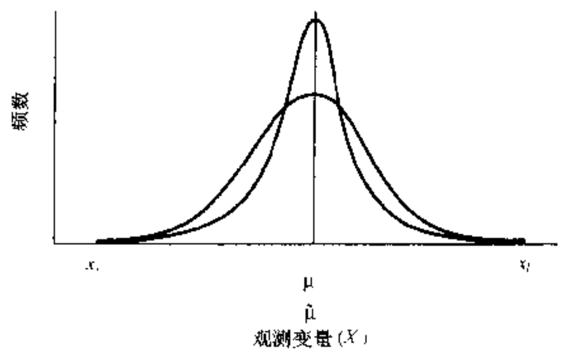


图 7-1

我们考虑两种离散性度量：极差(见 1.1 节)和离均差(见 6.4 节).尽管离均差在本章提出的离散性度量中起着重要作用，由图 7-1 仍可清楚地看到极差作为离散性的一种度量具有有限值.虽然两条曲线的离散程度明显不同，但它们有相同的极差 $x_l - x_s$.当然极差对于确定数据散布的外在界限是很重要的，但对于这两个极限间的分布状况，并未给出任何信息；此外，极差不大可靠，因为它对极值非常敏感.

7.2　平均偏差

我们在第六章给出了几种集中趋势的度量，其中算术平均是目前最重要也是最常用的一种度量，因而我们需要给出一种刻画相对于均值的离散程度的度量.(从本章起，除非有特殊说明，**术语"均值"均指"算术平均"**).这种度量的理想形式应为：(1)利用全部数据计算得到；(2)仅用一个数值刻画了数据相对于均值的平均偏离程度；(3)对于不同的数据集，当离散程度变大时，该数值亦变大.

对于任意一个样本，离散程度的一种显然的度量是与均值的偏离 $x_i - \bar{x}$(见 7.1 节)，从而相对于均值的一种平均偏离度量是这些偏差的算术平均

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n} \tag{7.1}$$

这个计算公式显然满足准则(1)(2)，即可由所有的样本数据算出，而且刻画了相对于均值的平均偏离程度，但它不满足准则(3).不论数据的离散程度如何，由该计算公式得到的结果恒为零.正如我们在 6.4 节中证明的那样，$\sum_{i=1}^{n}(x_i - \bar{x})$ 恒为零.

上述问题来源于均值的一条基本性质：它是**分布的重心**(见 6.4 节).正因为如此，相对于均值的正偏差的和恒等于负偏差的和，这两部分再求和时结果为零.为解决这个问题，应在计

算中消除负号的影响,我们有两种方法.其一是采用下式

$$\frac{\sum_{i=1}^{n}|x_i-\bar{x}|}{n} \tag{7.2}$$

这里分子是偏差的**绝对值**的和(见1.5节),而绝对值总为正的;另外一种方法是采用**方差**及**标准差**(见7.5和7.9节),即将每个偏差平方然后在计算中用**偏差的平方和**.

我们称(7.2)式为**平均偏差(平均绝对偏差)**.它刻画了数据对于均值的平均偏离程度,不考虑偏离的方向.当样本中的所有值都相同时,其平均偏差为零,当离散程度变大时,平均偏差亦增大,尽管平均偏差是离均差的一种可行的度量,但由于它在理论统计中的价值有限,故很少使用.

总体的另外一种平均偏差的计算公式为

$$\frac{\sum_{i=1}^{N}|x_i-\mu|}{N} \tag{7.3}$$

这两个公式有时修整为关于中位数的平均偏离

$$\frac{\sum_{i=1}^{n}|x_i-\bar{x}|}{n} \tag{7.4}$$

或

$$\frac{\sum_{i=1}^{N}|x_i-\bar{\mu}|}{N} \tag{7.5}$$

例 7.1　计算例6.1(a),(b)中样本的极差和平均偏差(关于舍入的讨论见6.3节)

解　(a) 极差$=x_l-x_s=7\text{g}$(克)$-1\text{g}=6\text{g}$

$$\bar{x}=\frac{\sum_{i=1}^{7}x_i}{7}=\frac{24\text{g}}{7}=3.42857\text{g}$$

$$\begin{aligned}\sum_{i=1}^{7}|x_i-\bar{x}|&=|1-3.42857|\,\text{g}+|3-3.42857|\,\text{g}+|2-3.42857|\,\text{g}+|7-3.42857|\,\text{g}+\\&\quad|5-3.42857|\,\text{g}+|4-3.42857|\,\text{g}+|2-3.42857|\,\text{g}\\&=11.42857\ \text{g}\end{aligned}$$

$$\text{平均偏差}=\frac{\sum_{i=1}^{7}|x_i-\bar{x}|}{7}=\frac{11.42857\ \text{g}}{7}=1.63265\ \text{g}\approx 1.6\ \text{g}$$

(b) 极差$=x_l-x_s=200\ \text{g}-1\ \text{g}=199\ \text{g}$

$$\bar{x}=\frac{\sum_{i=1}^{7}x_i}{7}=\frac{222\ \text{g}}{7}=31.7143\ \text{g}$$

$$\begin{aligned}\sum_{i=1}^{7}|x_i-\bar{x}|&=|1-31.7143|\,\text{g}+|3-31.7143|\,\text{g}+|2-31.7143|\,\text{g}+|7-31.7143|\,\text{g}\\&\quad+|5-31.7143|\,\text{g}+|4-31.7143|\,\text{g}+|200-31.7143|\,\text{g}\\&=336.5715\ \text{g}\end{aligned}$$

$$\text{平均偏差}=\frac{\sum_{i=1}^{7}|x_i-\bar{x}|}{7}=\frac{336.5715\ \text{g}}{7}=48.0816\ \text{g}\approx 48.1\ \text{g}$$

7.3　平均偏差的频数分布式

对于样本均值 $\bar{x}$,我们有频数分布公式(见6.6节)(类似于总体平均偏差公式),同样我

们也有**样本平均偏差的频数分布公式**

$$\text{平均偏差}=\frac{\sum_{i=1}^{k} f_i \mid x_i-\bar{x} \mid}{\sum_{i=1}^{k} f_i} \tag{7.6}$$

$$=\frac{\sum_{i=1}^{k} f_i \mid x_i-\bar{x} \mid}{n} \tag{7.7}$$

例 7.2　计算表 6.1 中样本的极差和平均偏差

解 ☞　极差=1.8 厘米 -1.2 厘米=0.6 厘米

为计算平均偏差,需要在表 6.1 中添加两列:$x_i-\bar{x}$ 和 $f_i|x_i-\bar{x}|$. 这两列及相应的平均偏差计算结果见表 7.1.

表 7.1

长度(厘米)x_i	频数 f_i	f_ix_i(厘米)	$(x_i-\bar{x})$(厘米)	$f_i\|x_i-\bar{x}\|$(厘米)
1.2	2	2.4	−0.30	0.60
1.3	7	9.1	−0.20	1.40
1.4	10	14.0	−0.10	1.00
1.5	12	18.0	0.00	0.00
1.6	10	16.0	0.10	1.00
1.7	7	11.9	0.20	1.40
1.8	2	3.6	0.30	0.60
$\sum$	50	75.0 厘米		6.00 厘米

$$\bar{x}=\frac{\sum f_ix_i}{n}=\frac{75.0\text{ 厘米}}{50}=1.50\text{ 厘米}$$

$$\text{平均偏差}=\frac{\sum f_i \mid x_i-\bar{x} \mid}{n}=\frac{6.00\text{ 厘米}}{50}=0.12\text{ 厘米}$$

7.4　近似平均偏差

由于通过分组频数分布得到的平均偏差只是由样本数据直接计算的平均偏差的精确值的近似,故称之为**近似平均偏差**. 正如我们对近似算术平均所做的假定一样(见 6.7 节),这里对分组数据也必须假定在组内的所有值都等于组标 m_i. 然而,对于某个总体,其近似平均偏差可如下计算

$$\text{总体平均偏差}\approx\frac{\sum_{i=1}^{k} f_i \mid m_i-(\approx\mu) \mid}{N} \tag{7.8}$$

对于一个样本,可如下计算

$$\text{样本平均偏差}\approx\frac{\sum_{i=1}^{k} f_i \mid m_i-(\approx\bar{x}) \mid}{n} \tag{7.9}$$

这里符号$(\approx\mu)$与$(\approx\bar{x})$表示近似等于均值.

例 7.3　计算表 6.3 中分组分布的 30 个马拉松赛跑时间的近似极差和近似平均偏差.

解 ☞　由习题 6.28,对于分组数据,我们有两种方法来近似得到 x_s 和 x_l. 从而,由方法(1)可得近似极差为

$$极差 \approx 144\text{分钟} - 129\text{分钟} = 15\text{分钟}$$

由方法(2)可得近似极差为

$$极差 \approx 145.5\text{分钟} - 127.5\text{分钟} = 18.0\text{分钟}$$

对于表 6.3 中的数据,利用(7.8)计算总体的近似平均偏差.需要在该表中添加另外两列数据:$m_i - (\approx \mu)$和 $f_i|m_i - (\approx\mu)|$.所需结果及总体的近似平均偏差结果见表 7.2.

表 7.2

时间(分钟)	组标 m_i(分钟)	频数 f_i	$f_i m_i$(分钟)	$[m_i-(\approx\mu)]$(分钟)	$f_i\|m_i-(\approx\mu)\|$(分钟)
128—130	129	3	387	-10.6	31.8
131—133	132	1	132	-7.6	7.6
134—136	135	4	540	-4.6	18.4
137—139	138	3	414	-1.6	4.8
140—142	141	7	987	1.4	9.8
143—145	144	12	1,728	4.4	52.8
$\sum$		30	4,188 分钟		125.2 分钟

$$\mu \approx \frac{\sum f_i m_i}{N} = \frac{4\,188\text{分钟}}{30} = 139.6\text{分钟}$$

$$总体平均偏差 \approx \frac{\sum f_i \mid m_i - (\approx\mu) \mid}{N} = \frac{125.2\text{分钟}}{30} = 4.17333\text{分钟} \approx 4.2\text{分钟}$$

7.5 总体方差:定义式

我们在 7.2 节中指出,度量某一数据集相对于其均值的偏离程度,我们将所有的与均值的偏差平方,然后求和,这个和式称之为**平方和**(记为 SS),样本平方和为

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{7.10}$$

总体平方和为

$$\sum_{i=1}^{N}(x_i - \mu)^2 \tag{7.11}$$

总体的**方差(均方差,均方和)**是这些关于总体均值的偏差的平方的算术平均.于是我们有**总体方差的定义式**

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} \tag{7.12}$$

或

$$\sigma^2 = \frac{SS}{N} \tag{7.13}$$

这里σ^2表示总体方差,σ是小写希腊字母(1.22 节中 $\sum$ 是其大写字母)

例 7.4 利用(7.12)式计算重量总体(单位:克):2, 3, 4, 5, 6 的方差.

解 ☞

$$\mu = \frac{\sum_{i=1}^{N}x_i}{N} = \frac{2\text{克}+3\text{克}+4\text{克}+5\text{克}+6\text{克}}{5} = \frac{20\text{克}}{5} = 4\text{克}$$

因此

$$\sigma^2 = \frac{\sum_{i=1}^{N}x_i - \mu}{N} = \frac{(2\text{克}-4\text{克})^2+(3\text{克}-4\text{克})^2+(4\text{克}-4\text{克})^2+(5\text{克}-4\text{克})^2+(6\text{克}-4\text{克})^2}{5}$$

$$=\frac{10\text{ 克}^2}{5}=2\text{ 克}^2$$

注：方差的单位是原测量单位的平方.

7.6　总体方差：计算式

总体方差的定义式有一个缺点：需要从每个数据减去均值.为消除这个缺点，我们给出它的另外一个等价式——计算式

$$\sigma^2=\frac{\sum_{i=1}^{N}x_i^2-\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}}{N} \tag{7.14}$$

及

$$\sigma^2=\frac{\sum_{i=1}^{N}x_i^2}{N}-\mu^2 \tag{7.15}$$

推导过程见习题 7.5，7.6.

由于上述表达式在计算中更简便易用，故称之为**计算式（机器式或工作式）**.

例 7.5　利用(7.14)式求例 7.4 中重量总体的方差.

解 ☞　由例 7.4 知 $N=5$，$\sum x_i=20$ 克，故

$$\sum_{i=1}^{N}x_i^2=4\text{ 克}^2+9\text{ 克}^2+16\text{ 克}^2+25\text{ 克}^2+36\text{ 克}^2=90\text{ 克}^2$$

因而

$$\sigma^2=\frac{\sum_{i=1}^{N}x_i^2-\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}}{N}=\frac{90\text{ 克}^2-\frac{(20\text{ 克})^2}{5}}{5}=2\text{ 克}^2$$

7.7　样本方差：定义式

对于一个样本，其方差定义如下

$$\text{样本方差}=s^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \tag{7.16}$$

分子是样本平方和 SS[见(7.10)式].总体方差的分母为 N(总体的大小)[见(7.12)式]，而这里分母为 $n-1$，样本容量 n 减 1.那么为什么是 $n-1$，而不是 n 呢？

为理解这个问题，我们必须回到 3.14 节中所讲过的内容，在那里，我们提到了两种水平的统计量：一种是数学水平的，我们推导并证明了一系列完整的数学统计量；另外一种是一般统计的，这主要是一系列统计概念和方法的非数学讨论.需要注意的是，本书提到的概念均是在数学水平上，不加证明而承认其为真，这里就是一例.

样本描述性度量(称为统计量)是用来估计总体的描述性度量(称为参数)(见 3.4 节).这个估计量必须满足几条性质，包括无偏性(没有系统的过高或过低估计参数).从数学上可以证明，样本均值 $\bar{x}$ 是总体均值 μ 的一个无偏估计，还可以证明，偏差平方的算术平均 $\frac{\sum(x_i-\bar{x})^2}{n}$ 是总体方差 σ^2 的一个有偏估计，它低估了 σ^2.此外还可证明，若将样本偏差平方和除以 $n-1$，则可消除有偏性.

例 7.6　利用(7.16)式计算关于长度的样本(单位：cm(厘米))：3，4，5，6，7 的方差.

解 ☞

$$\bar{x}=\frac{\sum x_i}{n}=\frac{25\text{cm}}{5}=5\text{cm}$$

因而

$$\begin{aligned}s^2&=\frac{\sum(x_i-\bar{x})^2}{n-1}\\&=\frac{(3\text{cm}-5\text{cm})^2+(4\text{cm}-5\text{cm})^2+(5\text{cm}-5\text{cm})^2+(6\text{cm}-5\text{cm})^2+(7\text{cm}-5\text{cm})^2}{5-1}\\&=\frac{10\text{cm}^2}{4}=2.5\text{cm}^2\end{aligned}$$

7.8 样本方差:计算式

总体方差有定义式和计算式,二者等价(见 7.6 节).样本方差同样有定义式和计算式,且二者等价,这里有三种计算式(推导过程见习题 7.8)

$$s^2=\frac{\sum_{i=1}^{n}x_i{}^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}{n-1}\tag{7.17}$$

$$s^2=\frac{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}{n(n-1)}\tag{7.18}$$

$$s^2=\frac{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}{n-1}\tag{7.19}$$

例 7.7 利用(7.17)式计算例 7.6 中关于长度的样本方差.

解 ☞ 由例 7.6 知 $n=5$, $\sum x_i=25\text{ cm}$,故

$$\sum x_i^2=9\text{cm}^2+16\text{cm}^2+25\text{cm}^2+36\text{cm}^2+49\text{cm}^2=135\text{cm}^2$$

因而

$$s^2=\frac{\sum x_i^2-\frac{(\sum x_i)^2}{n}}{n-1}=\frac{135\text{cm}^2-\frac{(25\text{cm})^2}{5}}{5-1}=2.5\text{cm}^2$$

7.9 总体标准差

作为数据关于均值的离散程度的一种度量,方差满足 7.2 节给出的三个准则:利用所有数据,仅用一个数值(偏差的平方)刻画了数据相对于均值的离散程度,当离散程度增大时,该数值亦增大.然而,它亦有两个重要的局限性:(1)方差的单位是原始观测数据的单位的平方,然而刻画离散程度的一种理想度量应该具有与均值相同的单位;(2)由于方差与平方有关,故它对数据集中的极值过于敏感.解决这两个局限性的一种方法是取方差的正的平方根,我们称该数值为**标准差**(或**根均方偏差**),标准差的单位则与原始测量单位相同,且对极值的敏感程度较弱.

总体标准差有下述定义式(见 7.5 节)

$$\sigma=\sqrt{\sigma^2}\tag{7.20}$$

$$\sigma=\sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}\tag{7.21}$$

其计算式为(见 7.6 节)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N}} \tag{7.22}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i^2)}{N} - \mu^2} \tag{7.23}$$

例 7.8　利用(7.20)式计算例 7.4 中总体方差的标准差.

解 ☞

$$\sigma = \sqrt{\sigma^2} = \sqrt{2\text{g}^2} = 1.41421\text{g} \approx 1.4\text{g}$$

7.10　样本标准差

样本标准差如下定义(见 7.7 节)

$$s = \sqrt{s^2} \tag{7.24}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \tag{7.25}$$

其计算式为(见 7.8 节)

$$s^2 = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}} \tag{7.26}$$

$$s^2 = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}} \tag{7.27}$$

$$s^2 = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}} \tag{7.28}$$

总体标准差和样本标准差**恒为正值**,因为它们是相应方差的正的平方根(或主平方根,见 1.7 节).

尽管方差在统计推断中有很多用处,在本书中我们也多次用到方差,但由于它具有我们刚刚讨论过的一些局限性,使得它很少作为一种描述性度量.相反,在描述性统计和推断统计中,标准差更常用来作为离均差的度量.

例 7.9　利用(7.24)式求例 7.6 中样本标准差.

解 ☞

$$s = \sqrt{s^2} = \sqrt{2.5\text{cm}^2} = 1.58114\text{cm} \approx 1.6\text{cm}$$

7.11　离散度量中的舍入原则

在 6.3 节,我们曾谈到,对于算术平均,在计算这种度量或使用这种度量时,应坚持最小的舍入;在计算中,至少应保留 6 个有效数字,在计算或使用离散性度量时亦如此.

在 6.3 节我们还谈到,对于不同问题最终结果的给出,虽然没有统一的绝对准则,但存在一个共同接受的原则:

如果所有数据的精度相同(见 2.15 节),那么均值应保留到精度的下一位.

将这一原则推广到离散性度量,有:

如果所有数据有相同的精度,则平均偏差及标准应保留到该精度水平的下一位,而方差应保留到该精度水平的下两位.

在本章对结果舍入时,一直采用这个**精度水平准则**.

最后,在 6.3 节我们谈到给出均值时还存在另一个常用的原则,即应与标准差协调,这个新的原则,称为**标准差原则**,内容如下:

将标准差舍入到两位有效数字,然后将均值舍入到标准差的最后一个数字位置,方差的小数位可以是标准差小数位的两倍,平均偏差也应当舍入到两位有效数字.

这两个舍入原则是等效的,以后本书均用它对结果进行舍入,此外,在第 13 章,我们将引入**标准误差原则**.

例 7.10 对于样本:4.9, 5.2, 6.1, 5.8, 7.3, 8.2, 6.5; $\bar{x}=6.28571$,平均偏差=0.89796, $s^2=1.35149$, $s=1.16254$. 分别采用如下原则,对这些值进行舍入:(a)精度水平原则,(b)标准差原则.

解 ☞ (a) $\bar{x}=6.29$, 平均偏差=0.90, $s^2=1.351$, $s=1.16$.

(b) $s=1.2$, $\bar{x}=6.3$, $s^2=1.35$,平均偏差=0.90.

7.12 由非分组频数分布计算标准差

为利用频数分布,应将总体标准差的公式进行变换(见 7.9 节),**总体标准差的频数分布定义式**为

$$\sigma=\sqrt{\frac{\sum_{i=1}^{k} f_i(x_i-\mu)^2}{N}} \tag{7.29}$$

总体标准差的频数分布计算式为

$$\sigma=\sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2-\frac{\left(\sum_{i=1}^{k} f_i x_i\right)^2}{N}}{N}} \tag{7.30}$$

$$\sigma=\sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2}{N}-\mu^2} \tag{7.31}$$

为利用频数分布,应将样本标准差的公式进行变换(见 7.10 节),**样本标准差的频数分布定义式**为

$$s=\sqrt{\frac{\sum_{i=1}^{k} f_i(x_i-\bar{x})^2}{n-1}} \tag{7.32}$$

样本标准差的频数分布计算式为

$$s=\sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2-\frac{\left(\sum_{i=1}^{k} f_i x_i\right)^2}{n}}{n-1}} \tag{7.33}$$

$$s=\sqrt{\frac{n\sum_{i=1}^{k} f_i x_i^2-\left(\sum_{i=1}^{k} f_i x_i\right)^2}{n(n-1)}} \tag{7.34}$$

$$s=\sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2-n\bar{x}^2}{n-1}} \tag{7.35}$$

例 7.11　利用(7.29)式求表 6.2 给出的马拉松赛跑时间的总体标准差.

解 ☞　若利用(7.29)式,需在表 6.2 中添加三列:$x_i-\mu$, $(x_i-\mu)^2$, $f_i(x_i-\mu)^2$,修改后的表及标准差的计算结果见表 7.3.

表 7.3

时间(分钟)x_i	频数 f_i	f_ix_i(分钟)	$(x_i-\mu)$(分钟)	$(x_i-\mu)^2$(分钟^2)	$f_i(x_i-\mu)^2$(分钟^2)
129	1	129	−10.8667	118.0852	118.0852
130	2	260	−9.8667	97.3518	194.7036
131	0	0	−8.8667	78.6184	0
132	0	0	−7.8667	61.8850	0
133	1	133	−6.8667	47.1516	47.1516
134	1	134	−5.8667	34.4182	34.4182
135	1	135	−4.8667	23.6848	23.6848
136	2	272	−3.8667	14.9514	29.9028
137	0	0	−2.8667	8.2180	0
138	3	414	−1.8667	3.4846	10.4538
139	0	0	−0.8667	0.7512	0
140	0	0	0.1333	0.0178	0
141	3	423	1.1333	1.2844	3.8532
142	4	568	2.1333	4.5510	18.2040
143	5	715	3.1333	9.8176	49.0880
144	2	288	4.1333	17.0842	34.1684
145	5	725	5.1333	26.3508	131.7540
$\sum$	30	4,196 分钟			695.4676 分钟^2

$$\mu=\frac{\sum f_ix_i}{N}=\frac{4,196\text{ 分钟}}{30}=139.8667\text{ 分钟}\approx 139.9\text{ 分钟}$$

$$\sigma=\sqrt{\frac{f_i(x_i-\mu)^2}{N}}=\sqrt{\frac{695.4676\text{ 分钟}^2}{30}}=4.8148\text{ 分钟}\approx 4.8\text{ 分钟}$$

7.13　由分组频数分布计算近似标准差

由分组频数分布求得的标准差仅近似于直接由数据求得的标准差,从而称之为**近似标准差**. 若想利用分组数据求得标准差,需假定某组内的所有值都等于其组标 m_i,然后利用下述**总体的计算式**求得总体的近似标准差

$$\sigma\approx\sqrt{\frac{\sum_{i=1}^{k}f_im_i^2-\frac{\left(\sum_{i=1}^{k}f_im_i\right)^2}{N}}{N}} \tag{7.36}$$

或利用下述**样本的计算式**求得样本的近似标准差

$$s\approx\sqrt{\frac{n\sum_{i=1}^{k}f_im_i^2-\left(\sum_{i=1}^{k}f_im_i\right)^2}{n(n-1)}} \tag{7.37}$$

例 7.12　利用(7.36)式,由表 6.3 中的马拉松赛跑时间的分组频数分布计算近似总体标准差.

解 ☞　若利用(7.36)式,需在表 6.3 中添加两列:m_i^2, $f_i m_i^2$. 修改后的表及近似总体标准差的计算结果见表 7.4.

表 7.4

时间(分钟)	组标 m_i(分钟)	频数 f_i	$f_i m_i$(分钟)	m_i^2(分钟²)	$f_i m_i^2$(分钟²)
128—130	129	3	387	16,641	49,923
131—133	132	1	132	17,424	17,424
134—136	135	4	540	18,225	72,900
137—139	138	3	414	19,044	57,132
140—142	141	7	987	19,881	139,167
143—145	144	12	1,728	20,736	248,832
$\sum$		30	4,188 分钟		585,378 分钟²

$$\sigma \approx \sqrt{\frac{\sum f_i m_i^2 - \frac{(\sum f_i m_i)^2}{N}}{N}} = \sqrt{\frac{585,378\ 分钟^2 - \frac{(1,188\ 分钟)^2}{30}}{30}}$$

$$= 4.94368\ 分钟 \approx 4.9\ 分钟$$

7.14　计算编码数据的方差和标准差

在 6.8 节,我们利用编码公式 $c_i = a - bx_i$[即(6.14)式]来计算样本算术平均. 现在我们利用该表达式计算样本方差和样本标准差,6.8 节中的一些术语含义在此处继续使用.

对样本进行编码后,利用(7.16)式可得 c_i 的方差

$$s_c^2 = \frac{\sum_{i=1}^{n}(c_i - \overline{c})^2}{n-1} \tag{7.38}$$

或等价地由(7.18)式

$$s_c^2 = \frac{n\sum_{i=1}^{n}c_i^2 - \left(\sum_{i=1}^{n}c_i\right)^2}{n(n-1)} \tag{7.39}$$

则原始数据的方差可由下述**方差解码公式**求得

$$s_x^2 = \frac{s_c^2}{b^2} \tag{7.40}$$

从而**标准差的解码公式**为

$$\sqrt{s_x^2} = \sqrt{\frac{s_c^2}{b^2}}$$
$$s_x = \frac{s_c}{b} \tag{7.41}$$

例 7.13　考虑表 6.4 中长度测量值的样本(单位:厘米),首先利用(7.18)和(7.24)式直接计算 s_x^2 和 s_x,然后利用(6.14),(7.39),(7.40)和(7.41)式重新计算 S_x^2 和 S_x. 在编码公式中令 $a=-490$ 厘米,$b=1$ 作为编码常量.

解 ☞　计算结果见表 7.5.

表 7.5

长度(厘米)x_i	x_i^2(厘米²)	($c_i = x_i - 490$)(厘米)	c_i^2(厘米²)
492	242,064	2	4
493	243,049	3	9
495	245,025	5	25
496	246,016	6	36
498	248,004	8	64
500	250,000	10	100
$\sum$ 2,974 厘米	1,474,158 厘米²	34 厘米	238 厘米²

续表 7.5

直接计算:

$$s_x^2=\frac{n\sum x_i^2-(\sum x_i)^2}{n(n-1)}=\frac{6(1,474,158\text{ 厘米}^2)-(2,974\text{ 厘米})^2}{6(5)}$$

$$=9.06667\text{ 厘米}^2\approx 9.07\text{ 厘米}^2$$

$$s_x=\sqrt{s_x^2}=\sqrt{9.06667\text{ 厘米}^2}=3.01109\text{ 厘米}\approx 3.0\text{ 厘米}$$

利用编码和解码公式计算:

$$s_c^2=\frac{n\sum c_i^2-(\sum c_i)^2}{n(n-1)}=\frac{6(238\text{ 厘米}^2)-(34\text{ 厘米})^2}{6(5)}=9.06667\text{ 厘米}^2$$

$$s_c=\sqrt{s_c^2}=\sqrt{9.06667\text{ 厘米}^2}=3.01109\text{ 厘米}$$

对 s_c^2 解码: $s_x^2=\frac{s_c^2}{b^2}=\frac{9.06667\text{ 厘米}^2}{1^2}=9.06667\text{ 厘米}^2\approx 9.07\text{ 厘米}^2$

对 s_c 解码: $s_x=\frac{s_c}{b}=\frac{3.01109\text{ 厘米}}{1}=3.01109\text{ 厘米}\approx 3.0\text{ 厘米}$

7.15　Chebyshev 定理

Chebyshev 定理:给定一组数据 $x_1, x_2, \cdots, x_n$(或 $x_1\ x_2, \cdots, x_N$),其均值假定为 m,标准差为 $\sigma>0$,则对于任意 $k\geqslant 1$,位于区间$[m-k\sigma, m+k\sigma]$内的数据所占比例大于等于 $1-\frac{1}{k^2}$.

图 7-2 给出了 Chebyshev 定理的一种解释. 对于给定的总体,位于区间$[\mu-k\sigma,\mu+k\sigma]$内的总体比例至少为 $1-\frac{1}{k^2}$(即曲线下的阴影部分). 这是总体比例的下限(或下界),明显地,位于区间 $\mu\pm k\sigma$(或 $\bar{x}\pm ks$)内的比例大于 $1-\frac{1}{k^2}$. 这个定理亦称为 **Chebyshev 不等式**(见 1.23 节),因为它实际上表明比例大于或等于($\geqslant$)$1-\frac{1}{k^2}$. 可以从数学上证明,对于任意的测量值集合,Chebyshev 定理都成立,不论是对称分布,有偏斜的分布或者多峰分布.

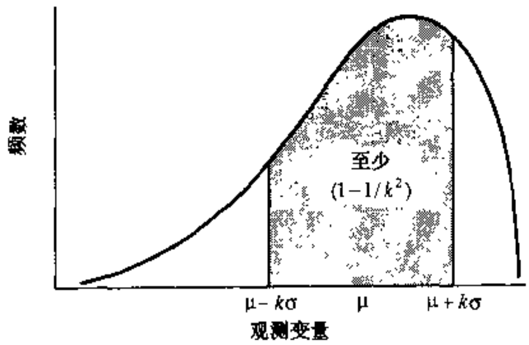


图 7-2

表 7.6

与 μ 相距 σ 或 s 的倍数 k	位于区间 $\mu\pm k\sigma$ 内的最少量	
	比例 $1-1/k^2$	百分比 $[(1-1/k^2)\times(100)](\%)$
1	0	00.0
$1\frac{1}{2}$		
2		
$2\frac{1}{2}$		
3		

表 7.7

与 μ 相距 σ 或 s 的倍数 k	位于区间 $\mu\pm k\sigma$ 内的最少量	
	比例 $1-1/k^2$	百分比 $[(1-1/k^2)\times(100)](\%)$
1	0	00.0
$1\frac{1}{2}$	5/9=0.556	55.6
2	3/4=0.750	75.0
$2\frac{1}{2}$	21/25=0.840	84.0
3	8/9=0.889	88.9

例 7.14　利用 Chebyshev 定理完成表 7.6

解　表 7.7 即为所求的表.

注:尽管对于任意 $k \geq 1$,Chebyshev 定理均成立,由表可知该定理仅在 $k>1$ 时,才能提供有意义的信息.

7.16　经验法则

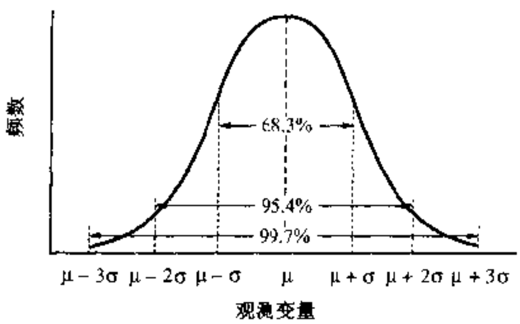


图 7-3

正态分布是一条理论上的钟形曲线,由**正态概率密度函数**(见第 12 章)定义.给定一组数据,如果它的频数折线或相对频数折线可由正态分布精确拟合,则称这组数据服从正态分布,此时,位于区间"均值± k 个标准差"内的数据所占百分比可由正态概率密度函数在该区间内的积分求得.图 7-3 给出了服从某正态分布的数据落入 $\mu\pm\sigma$, $\mu\pm2\sigma$ 和 $\mu\pm3\sigma$ 内的百分比,这些百分比对于任意服从正态分布的多边形均成立.

这些百分比对于那些近似服从正态分布的数据集也近似成立(如单峰,略呈圆顶状,基本对称的).由于这种推广是凭经验而得的,故称之为**经验法则**,具体陈述如下:

对于任意近似正态分布的总体,大约 68%的数据位于区间 $\mu\pm\sigma$ 内,大约 95%的数据位于区间 $\mu\pm2\sigma$ 内,大约 100%的数据位于区间 $\mu\pm3\sigma$ 内.

这种推广对于近似正态分布的样本亦成立.

经验法则和 Chebyshev 定理都属于描述性统计,因为在给定某一分布及其均值和标准差时,这两种方法可以迅速地计算出距离均值一定范围内的数据所占的百分比.如果分布至少渐近正态分布,那末经验法则给出了距离均值一个,二个或三个标准差的数据所占的近似百分比.如果分布未知或与正态分布差别较大,则 Chebyshev 定理可以算出距离均值 k 个标准差的数据所占百分比的下限(下界).

尽管这个法则在统计书中称为经验法则,但它不是我们使用的第一个经验决定法则.回忆习题 6.31,我们曾提到对于略有偏斜分布的"经验法则"

$$\bar{x}-\tilde{x}=\frac{1}{3}(\bar{x}-\text{众数})$$

例 7.15　若总体服从正态分布,求位于 $\mu-\sigma$ 左侧的数据所占的百分比

解　由图 7-3 知正态分布关于均值 μ 对称.如果位于区间 $\mu\pm\sigma$ 内的数据所占百分比为 68.3%,则位于区间 $\mu\pm\sigma$ 外的数据所占百分比为(100% －68.3%=31.7%).由于分布关于均值 μ 对称,故位于 $\mu-\sigma$ 左侧的数据所占百分比为

$$\frac{1}{2}(31.7\%)=15.85\%$$

7.17　集中趋势和偏离性的图示

演示某样本或总体集中趋势和偏离性的方法有三种,见图 7-4.由此可以得到 10 棵谷物植株在成熟后两周及四周时的平均高度(单位:英寸).在图 7-4(a)中,植株高度以条形图表示,其中 Y 轴表测量尺度,每个条形的高度表示在测量那天的植株平均高度 $\bar{x}$,每个条形上方的垂直线段长度,称为**误差条**,表示一个标准差 s,条形上面的这些线段可表示一个或多个标准差;或者均值的一个或多个标准误(见第 13 章),或者一个置信区间的一半(见第 14 章).

图 7-4(b)以另外一种方式刻画了同样的结果:穿过竖直矩形中心的水平线表示样本均

值 $\bar{x}$,这条直线到矩形的上底或下底的距离表示样本标准差,垂线表示样本极差(这种垂线在表示极差时,不能称为误差条).

最后,图 7-4(c)采用线图方式刻画了同样的结果,其中 Y 轴仍表示测量尺度,X 轴表示时间,X 轴上方每个实心圆点的高度表示在这一观测日的样本均值 $\bar{x}$,每个实心圆点上方和下方的垂直线(误差条)表示沿 $\bar{x}\pm s$ 方向的一个标准差,相邻的实心圆点用一直线段联结,这表明在连续时间内发生的事件是有内在联系的.

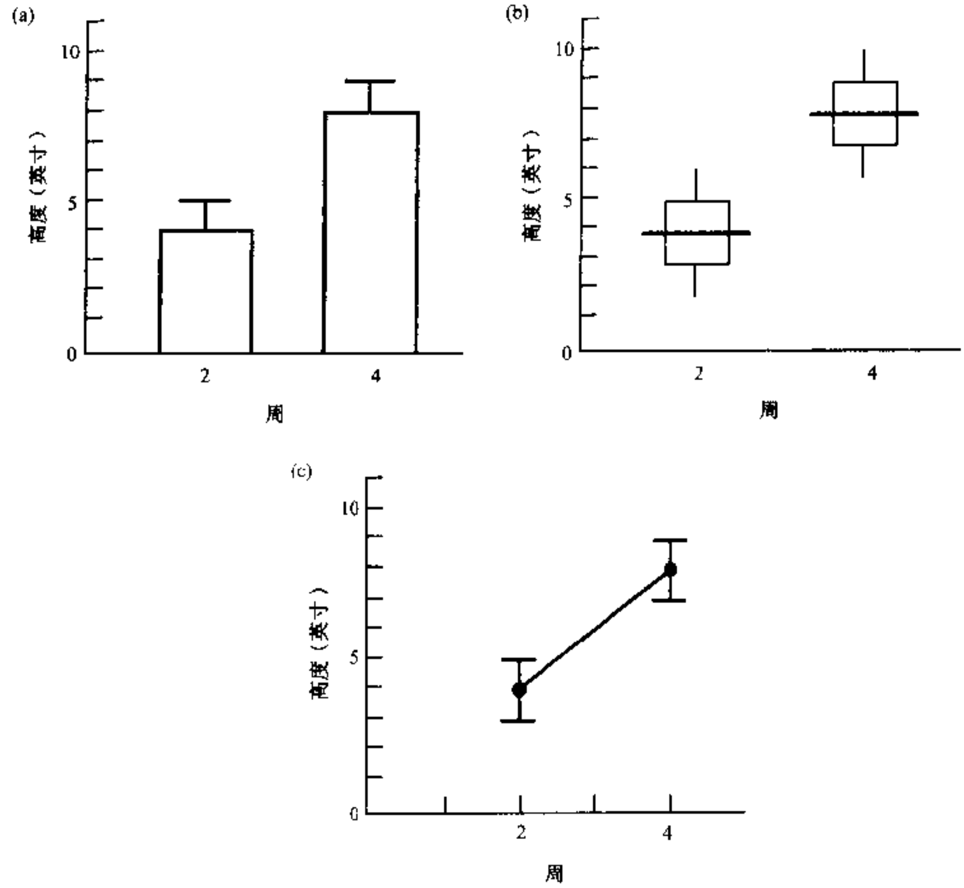


图 7-4

例 7.16　由图 7-4,求在第二周及第四周植株的平均高度,标准差及极差(精确到英寸)?

解　第二周:$\bar{x}=4$ 英寸,$s=1$ 英寸,极差=4 英寸;第四周:$\bar{x}=8$ 英寸,$s=1$ 英寸,极差=5 英寸.

7.18　变异系数

对于某一总体,我们定义**变异系数**(或**变差系数**,**离差系数**,**相对标准差**)为

$$V=\frac{\sigma}{\mu} \tag{7.42}$$

及

$$V=\frac{\sigma}{\mu}(100\%) \tag{7.43}$$

(7.42)式给出了标准差所占均值的比例,而(7.43)式则给出了标准差所占均值的百分比,我们更常用的是(7.43)式.类似地,对于某一样本,有

$$V=\frac{s}{\bar{x}} \tag{7.44}$$

及

$$V=\frac{s}{\bar{x}}(100\%) \tag{7.45}$$

前面我们所研究的偏离性度量(极差,平均偏差,方差,标准差)均称为**绝对偏离性度量**,因为它们均是由数据直接计算得到的,其单位或者与原始测量单位相同,或者是原始测量单位的平方.而变异系数可称为**相对偏离性度量**,因为它刻画了绝对偏离性度量所占均值的某种度量的比例(或百分比).由于分子,分母的单位相同,故变异系数是**没有单位**的相对偏离性度量.

例 7.17 一位生物学家研究不同种类的啮齿动物的遗传变异,对于每只啮齿动物,测量其体重(单位:克).他随机选取 10 只小白鼠,测得结果为:$\bar{x}=12.9$ 克,$s=1.6$ 克,又随机选取 8 只灰袋鼠,测得结果为:$\bar{x}=545.0$ 克,$s=32.8$ 克,试利用(7.45)式比较这两种动物体重的相对偏离程度.

解 ☞ 对于小白鼠

$$V=\frac{s}{\bar{x}}(100\%)=\frac{1.6\text{克}}{12.9\text{克}}(100\%)=12.4\%$$

对于灰袋鼠

$$V=\frac{s}{\bar{x}}(100\%)=\frac{32.8\text{克}}{545.0\text{克}}(100\%)=6.0\%$$

这些结果表明:小白鼠体重的相对偏离程度大约是灰袋鼠体重的相对偏离程度的两倍,这种相对于均值的较大变异性是无法由标准差体现的,标准差只是说明灰袋鼠的绝对变异程度大约是小白鼠的 20 倍.

注:一般地,对于几个样本(或总体),当均值增大时,其标准差往往也增大,因而比较几种具有不同均值的分布时,较大方差总是表明较大的绝对偏离程度,而变异系数则能更好地刻画相对偏离程度.这里袋鼠体重的均值大约是小白鼠体重的均值的 40 倍,这种方法还可用来比较小白鼠与更大哺乳动物的相对偏离程度,比如大象体重的均值大约为 7,000,000 克(大约是小白鼠体重均值的 540,000 倍).

7.19 标准分和标准化变量

总体的**标准分**(亦称为**正态偏差**,或 **z 分**)定义如下

$$z_i=\frac{x_i-\mu}{\sigma} \tag{7.46}$$

样本的标准分定义如下

$$z_i=\frac{x_i-\bar{x}}{s} \tag{7.47}$$

对于任意数据分布,标准分表明了任一给定的 x_i 与分布的均值以标准差为单位的距离,正的 z 值表明 x_i 大于均值(直方图或折线图中在它的左边),而 z 值为负表明 x_i 小于均值(在它的左边).与变异系数类似,标准分也是一种相对度量.变异系数刻画了数据相对于均值的绝对偏离程度,而标准分则刻画了数据与均值的相对于标准差的偏离程度.由于其单位是标准差的个数,因而标准分可用来比较具有不同均值或不同观测单位的分布之间的相对位置.

对于任意变量 X,将观测值转换成相应 z 值的过程称为将该变量**标准化**(或**正态化**),所得到的变量 Z 称为**标准化变量**.

例 7.18 考虑样本:3, 5, 7, 9, 11.首先求 $\bar{x}$ 和 s,然后将该样本标准化.

解 ☞

$$\bar{x}=\frac{\sum x_i}{n}=\frac{35}{5}=7$$

$$s=\sqrt{\frac{n\sum x_i^2-(\sum x_i)^2}{n(n-1)}}=\sqrt{\frac{5(285)-(35)^2}{5(4)}}=3.16228$$

为将此样本标准化,需对每个 x_i 计算标准分 z_i,结果如下(四舍五入到百分位):

$$z_1=\frac{x_1-\bar{x}}{s}=\frac{3-7}{3.16228}=-1.26491\approx-1.26$$

$$z_2 = \frac{x_2 - \bar{x}}{s} = \frac{5-7}{3.16228} = -0.63246 \approx -0.63$$

$$z_3 = \frac{x_3 - \bar{x}}{s} = \frac{7-7}{3.16228} = 0.00 \approx 0.00$$

$$z_4 = \frac{x_4 - \bar{x}}{s} = \frac{9-7}{3.16228} = 0.63246 \approx 0.63$$

$$z_5 = \frac{x_5 - \bar{x}}{s} = \frac{11-7}{3.16228} = 1.26491 \approx 1.26$$

7.20　四分位极差和四分位差

四分位极差是第一分位数与第三分位数的差,其定义如下

$$四分位极差 = Q_3 - Q_1 \tag{7.48}$$

这个区间包含分布中间的50%.

四分位差(或半四分位极差)是一种离散度量,其定义如下

$$四分位差 = \frac{Q_3 - Q_1}{2} \tag{7.49}$$

它是四分位极差的一半,从而包含中间数据的50%的区间的一半.

例 7.19　对例6.13中的四分位数(Q_1=1.385厘米, Q_3=1.615厘米),求:(a)四分位极差?(b)四分位差?

解　(a) 四分位极差$=Q_3-Q_1=$1.615厘米−1.385厘米=0.230厘米≈0.23厘米.

(b) 四分位差$=\frac{Q_3-Q_1}{2}=\frac{0.230厘米}{2}$=0.115厘米≈0.12厘米.

7.21　盒子图与五数概括

如果某分布是单峰且完全对称的,则可根据其均值及标准差并利用Chebyshev定理或者经验法则,对其作出很好的描述.然而,如果某分布是严重偏斜或者多峰的,则我们更常采用的是**五数概括**:Q_1, Q_2, Q_3, x_s, x_l.五数概括可以通过某种图形展示,称之为**盒子图**.

以图7-5为例演示这种盒子图.该盒子图是根据表6.1中的关于长度的数据及例6.13给出的四分位数而构成的.这种图可以沿x轴的测量尺度成水平形式(如这里),或沿y轴的测量尺度成竖直形式.盒子图的构造类似于图7-4(b)的漂浮矩形图的构造,只不过漂浮矩形展示的是$\bar{x}\pm s$与极差,而盒子图展示的是五数概括.这个矩形,称为**盒子**,沿观测尺度方向从Q_1延伸到Q_3,它的高度表示四分位极差,盒子的宽度可任意,它并未提供任何信息,穿过盒子的水平直线为$Q_2=\bar{x}$.正如图7-4(b)那样,盒子上方及下方的直线表示分布的极差,在盒子图中,这些直线称为**把**(whiskers).正如盒子图中所示那样,下面的把从x_s到Q_l,上面的把从Q_3到x_l(这里x_s是最小值,x_l是最大值).

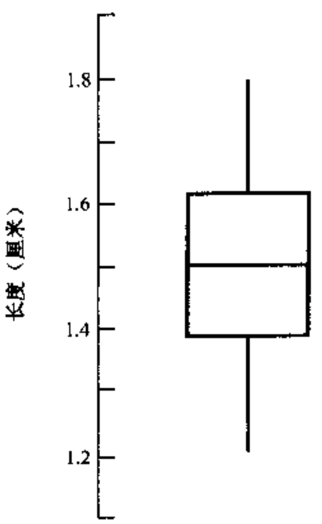


图7-5

尽管各种形式的盒子图均可表示Q_1, Q_2, Q_3,然而却有各种形式的把.

例如,在另外一种常用形式中,下方的把是从P_{10}到Q_1,上方的把是从Q_3到P_{90}. $P_{90}-P_{10}$被称为10—90**百分位极差**.当显示的是10—90百分位极差而不是极差时,就不能称之为五数概括.

习题解答

平均偏差

7.1 考虑习题 6.37 中的(a),(b),(c),(d),求样本的极差和平均偏差.

解 (a) 极差$=x_l-x_s=0$ 秒-0 秒$=0$ 秒

$$\bar{x}=\frac{\sum_{i=1}^{5}x_i}{5}=\frac{0\text{秒}}{5}=0.0\text{秒}$$

$$\sum_{i=1}^{5}|x_i-\bar{x}|=|0-0.0|\text{秒}+|0-0.0|\text{秒}+|0-0.0|\text{秒}+|0-0.0|\text{秒}+|0-0.0|\text{秒}$$
$$=0.0\text{秒}$$

$$\text{平均偏差}=\frac{\sum_{i=1}^{5}|x_i-\bar{x}|}{5}=\frac{0.0\text{秒}}{5}=0.0\text{秒}$$

(b) 极差$=x_l-x_s=10$ 秒-0 秒$=10$ 秒

$$\bar{x}=\frac{\sum_{i=1}^{5}x_i}{5}=\frac{10\text{秒}}{5}=2.0\text{秒}$$

$$\sum_{i=1}^{5}|x_i-\bar{x}|=|10-2.0|\text{秒}+|0-2.0|\text{秒}+|0-2.0|\text{秒}+|0-2.0|\text{秒}+|0-2.0|\text{秒}$$
$$=16.0\text{秒}$$

$$\text{平均偏差}=\frac{\sum_{i=1}^{5}|x_i-\bar{x}|}{5}=\frac{16.0\text{秒}}{5}=3.2\text{秒}$$

(c) 极差$=x_l-x_s=112.217$ km(公里)-9.777km$=102.440$km

$$\bar{x}=\frac{\sum_{i=1}^{6}x_i}{6}=\frac{172.459\text{km}}{65}=28.7432\text{km}$$

$$\sum_{i=1}^{6}|x_i-\bar{x}|=|10.127-28.7432|\text{km}+|11.963-28.7432|\text{km}+|112.217-28.7432|\text{km}$$
$$+|9.777-28.7432|\text{km}+|13.833-28.7432|\text{km}+|14.542-28.7432|\text{km}$$
$$=166.9478\text{km}$$

$$\text{平均偏差}=\frac{\sum_{i=1}^{6}|x_i-\bar{x}|}{6}=\frac{166.9478\text{ km}}{6}=27.8246\text{ km}$$

(d) 极差$=x_l-x_s=14.542$ km -0.007km$=14.535$km

$$\bar{x}=\frac{\sum_{i=1}^{6}x_i}{6}=\frac{60.249\text{ km}}{65}=10.0415\text{ km}$$

$$\sum_{i=1}^{6}|x_i-\bar{x}|=|10.127-10.0415|\text{km}+|11.963-10.0415|\text{km}+|112.217-10.0415|\text{km}+$$
$$|9.777-10.0415|\text{km}+|13.833-10.0415|\text{km}+|14.542-10.0415|\text{km}$$
$$=20.5980\text{ km}$$

$$\text{平均偏差}=\frac{\sum_{i=1}^{6}|x_i-\bar{x}|}{6}=\frac{20.5980\text{ km}}{6}=3.4330\text{ km}$$

7.2 考虑表 6.8 中的样本,计算极差,并利用(7.7)式计算其平均偏差.

解 ☞　极差=20 克−14 克=6 克

表 7.8

重量(克)x_i	频数 f_i	f_ix_i(克)	$(x_i-\bar{x})$(克)	$f_i\|x_i-\bar{x}\|$(克)
14	2	28	−4.0556	8.1112
15	2	30	−3.0556	6.1112
16	4	64	−2.0556	8.2224
17	18	306	−1.0556	19.0008
18	24	432	−0.0556	1.3344
19	35	665	0.9444	33.0540
20	5	100	1.9444	9.7220
$\sum$	90	1,625(克)		85.5560 克

$$\bar{x}=\frac{\sum f_ix_i}{n}=\frac{1,625\text{ 克}}{90}=18.0556\text{ 克}$$

$$\text{平均偏差}=\frac{\sum f_i\mid x_i-\bar{x}\mid}{n}=\frac{85.5560\text{ 克}}{90}=0.950622\text{ 克}\approx 1.0\text{ 克}$$

7.3　利用(7.9)式求表 6.10 中的分组数据的近似平均偏差.

解 ☞　修改后的表及计算结果见表 7.9.

表 7.9

第二次考试	组标 m_i	频数 f_i	f_im_i	$[m_i-(\approx\bar{x})]$	$f_i\|m_i-(\approx\bar{x})\|$
45—49	47	1	47	−34.375	34.375
50—54	52	0	0	−29.375	0
55—59	57	4	228	−24.375	97.500
60—64	62	3	186	−19.375	58.125
65—69	67	5	335	−14.375	71.875
70—74	72	3	216	−9.375	28.125
75—79	77	6	462	−4.375	26.250
80—84	82	11	902	0.625	6.875
85—89	87	9	783	5.625	50.625
90—94	92	17	1,564	10.625	180.625
95—99	97	5	485	15.625	78.125
$\sum$		64	5,208		632.500

$$\bar{x}\approx\frac{\sum f_im_i}{n}=\frac{5,208}{64}=81.375\approx 81.4$$

$$\text{样本平均偏差}\approx\frac{\sum f_i\mid m_i-(\approx\bar{x})\mid}{n}=\frac{632.500}{64}=9.88281\approx 9.9$$

总体方差

7.4　考虑例 7.4 中的重量总体,利用(7.15)式求总体方差.

解 ☞　由例 7.4 及例 7.5 知 $N=5$, $\mu=4$ 克, $\sum x_i^2=90$ 克^2.因而

$$\sigma^2=\frac{\sum_{i=1}^{N}x_i^2}{N}-\mu^2=\frac{90\text{ 克}^2}{5}-(4\text{ 克})^2=2\text{ 克}^2$$

7.5 证明

$$总体平方和 SS=\sum_{i=1}^{N}x_i^2-\frac{\left(\sum\limits_{i=1}^{N}x_i\right)^2}{N} \tag{7.50}$$

证明:在下列推导中,$\sum$ 表示 $\sum\limits_{i=1}^{N}$

$$总体 SS=\sum(x_i-\mu)^2=\sum(x_i^2-2x_i\mu+\mu^2)$$
$$=\sum x_i^2-\sum 2x_i\mu+\sum\mu^2$$

由于 2 及 μ 为常量,

$$总体 SS=\sum x_i^2-2\mu\sum x_i+N\mu^2$$
$$=\sum x_i^2-2\left(\frac{\sum x_i}{N}\right)\sum x_i+N\left(\frac{\sum x_i}{N}\right)^2$$
$$=\sum x_i^2-2\frac{\left(\sum x_i\right)^2}{N}+\frac{\left(\sum x_i\right)^2}{N}$$
$$=\sum x_i^2-\frac{\left(\sum x_i\right)^2}{N}$$

注:类似地,可以证明

$$样本平方和 SS=\sum_{i=1}^{n}x_i^2-\frac{\left(\sum\limits_{i=1}^{n}x_i\right)^2}{n} \tag{7.51}$$

7.6 由总体方差定义式[(7.12)式]推导总体方差计算式[(7.14)及(7.15)式].

解 为推导(7.14)式,将(7.50)式代入(7.12)式的分子.

为导出(7.15)式,需对(7.14)式做如下计算

$$\sigma^2=\frac{\sum x_i^2-\frac{\left(\sum x_i\right)^2}{N}}{N}$$
$$=\frac{1}{N}\left[\sum x_i^2-\frac{\left(\sum x_i\right)^2}{N}\right]$$
$$=\frac{\sum x_i^2}{N}-\frac{\left(\sum x_i\right)^2}{N^2}$$
$$=\frac{\sum x_i^2}{N}-\left(\frac{\sum x_i}{N}\right)^2$$
$$=\frac{\sum x_i^2}{N}-\mu^2$$

7.7 从数学上已证明:对于任意的 $x_1,x_2,\cdots,x_k$,平方和 $\sum\limits_{i=1}^{k}(x_i-a)^2$ 取最小值,当且仅当 $a=\frac{x_1+x_2+\cdots+x_k}{k}$. 对于样本:1, 2, 3, 4, 5 验证这一结论.

解 为验证当且仅当 $a=\bar{x}=\frac{\sum\limits_{i=1}^{5}x_i}{5}=\frac{15}{5}=3$ 时,$\sum\limits_{i=1}^{k}(x_i-a)^2$ 达到最小,我们需计算当 a 小于 $\bar{x}=3$ 或 a 大于 $\bar{x}=3$ 时,$\sum\limits_{i=1}^{5}(x_i-a)^2$ 的值. 这里取 $a=2.8, 2.9, 3.0, 3.1, 3.2$,计算结果见表 7.10. 由表的最后一行可见,当 $a\neq\bar{x}$ 时,$\sum(x_i-a)^2>\sum(x_i-\bar{x})^2$.

表 7.10

x_i	$(x_i-2.8)^2$	$(x_i-2.9)^2$	$(x_i-3.0)^2$	$(x_i-3.1)^2$	$(x_i-3.2)^2$
1	3.24	3.61	4	4.41	4.84
2	0.64	0.81	1	1.21	1.44
3	0.04	0.01	0	0.01	0.04
4	1.44	1.21	1	0.81	0.64
5	4.84	4.41	4	3.61	3.24
$\sum$	10.20	10.05	10	10.05	10.20

样本方差

7.8 由样本方差的定义式[(7.16)式]推导 7.8 节给出的样本方差的三个计算式[(7.17)，(7.18)，(7.19)式].

解 ☞ 为导出(7.17)式，将(7.51)式代入(7.16)式的分子中.

为导出(7.18)式，将(7.17)式两边同乘以$\frac{n}{n}$.

$$\frac{n}{n}s^2=\frac{n\left[\sum x_i^2-\frac{\left(\sum x_i\right)^2}{n}\right]}{n(n-1)}$$

$$s^2=\frac{n\sum x_i^2-\left(\sum x_i\right)^2}{n(n-1)}$$

最后，为导出(7.19)式，对(7.16)式做如下变形：

$$s^2=\frac{\sum(x_i-\bar{x})^2}{n-1}=\frac{\sum(x_i^2-2x_i\bar{x}+\bar{x}^2)}{n-1}=\frac{\sum x_i^2-\sum 2x_i\bar{x}+\sum\bar{x}^2}{n-1}$$

由于 2 和 $\bar{x}$ 均为常量，故

$$s^2=\frac{\sum x_i^2-2\bar{x}\sum x_i+n\bar{x}^2}{n-1}$$

又 $\sum x_i=n\bar{x}$ ，所以

$$s^2=\frac{\sum x_i^2-2\bar{x}(n\bar{x})+n\bar{x}^2}{n-1}=\frac{\sum x_i^2-2n\bar{x}^2+n\bar{x}^2}{n-1}=\frac{\sum x_i^2-n\bar{x}^2}{n-1}$$

7.9 用(7.18)及(7.19)式计算例 7.6 中长度的样本方差.

解 ☞ 由例 7.6，7.7 知，$n=5$，$\sum x_i=25$ 厘米，$\sum x_i^2=135$ 厘米，$\bar{x}=5$ 厘米. 因而，由(7.8)式得

$$s^2=\frac{n\sum x_i^2-\left(\sum x_i\right)^2}{n(n-1)}=\frac{5(135\text{ 厘米}^2)-(25\text{ 厘米})^2}{5(5-1)}=2.5\text{ 厘米}^2$$

由(7.19)式得

$$s^2=\frac{\sum x_i^2-n\bar{x}^2}{n-1}=\frac{135\text{ 厘米}^2-5(5\text{ 厘米})^2}{5-1}=2.5\text{ 厘米}^2$$

总体标准差

7.10 考虑总体：9.1，8.7，9.0，9.3，利用(7.21)式计算该总体的标准差.

解 ☞

$$\mu=\frac{\sum x_i}{N}=\frac{36.0}{4}=9.0$$

$$\sigma=\sqrt{\frac{\sum(x_i-\mu)^2}{N}}$$

$$=\sqrt{\frac{(9.1-9.0)^2+(8.7-9.0)^2+(9.0-9.0)^2+(9.2-9.0)^2}{4}}$$
$$=0.18708$$
$$\approx 0.19$$

7.11 考虑习题 7.10,利用(7.23)式计算该总体的标准差.

解 ☞ 由习题 7.10 知,$\mu=9.0$,而

$$\sum x_i^2=(9.1)^2+(8.7)^2+(9.0)^2+(9.2)^2=324.14$$

故

$$\sigma=\sqrt{\frac{\sum x_i^2}{N}-\mu^2}=\sqrt{\frac{324.14}{4}-(9.0)^2}=0.18708\approx 0.19$$

7.12 利用(7.30)式计算表 6.2 中给出的马拉松赛跑时间总体的标准差.

解 ☞ 若利用(7.30)式,需在表 6.2 中添加两列:x_i^2,$f_i x_i^2$. 修改后的表及标准差的计算结果见表 7.11.

表 7.11

时间(分钟)x_i	频数 f_i	$f_i x_i$(分钟)	x_i^2(分钟^2)	$f_i x_i^2$(分钟^2)
129	1	129	16,641	16,641
130	2	260	16,900	33,800
131	0	0	17,161	0
132	0	0	17,424	0
133	1	133	17,689	17,689
134	1	134	17,956	17,956
135	1	135	18,225	18,225
136	2	272	18,496	36,992
137	0	0	18,769	0
138	3	414	19,044	57,132
139	0	0	19,321	0
140	0	0	19,600	0
141	3	423	19,881	59,643
142	4	568	20,164	80,656
143	5	715	20,449	102,245
144	2	288	20,736	41,472
145	5	725	21,025	105,125
$\sum$	30	4,196 分钟		587,576 分钟^2

$$\sigma=\sqrt{\frac{\sum f_i x_i^2-\frac{(\sum f_i x_i)^2}{N}}{N}}$$

$$=\sqrt{\frac{587,576\text{ 分钟}^2-\frac{(4,196\text{ 分钟})^2}{30}}{30}}=4.81479\text{ 分钟}\approx 4.8\text{ 分钟}$$

样本标准差

7.13 考虑例 6.1(a)中的样本,利用(7.25)与(7.27)式计算其标准差.

解 ☞ 由(7.25)式,有

$$\bar{x}=\frac{\sum x_i}{n}=\frac{24\text{克}}{7}=3.42857\text{克}$$

$$s=\sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{25.7143\text{克}^2}{6}}=2.07020\text{克}\approx2.1\text{克}$$

由(7.27)式，有

$$\sum x_i=24\text{克}$$

$$\sum x_i^2=108\text{克}^2$$

$$s=\sqrt{\frac{n\sum x_i^2-\left(\sum x_i\right)^2}{n(n-1)}}=\sqrt{\frac{7(108\text{克}^2)-(24\text{克})^2}{7(6)}}=\sqrt{4.28571\text{克}^2}=2.07020\text{克}\approx2.1\text{克}$$

7.14　考虑例 6.1(b)中的样本，利用(7.25)与(7.28)式计算其标准差.

解 ☞　由(7.25)式，有

$$\bar{x}=\frac{\sum x_i}{n}=\frac{222\text{克}}{7}=31.7143\text{克}$$

$$s=\sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{33,063.429\text{克}^2}{6}}=74.2332\text{克}\approx74.2\text{克}$$

由(7.28)式，有

$$\sum x_i^2=40,104\text{克}^2$$

$$s=\sqrt{\frac{\sum x_i^2-n\bar{x}^2}{n-1}}=\sqrt{\frac{40,104\text{克}^2-7,040.578\text{克}^2}{6}}=74.2332\text{克}\approx74.2\text{克}$$

7.15　利用(7.32)式计算表 6.1 中长度样本的标准差.

解 ☞　为利用(7.32)式，需在表 6.1 中添加三列：$x_i-\bar{x}$，$(x_i-\bar{x})^2$，$f_i(x_i-\bar{x})^2$. 修改后的表及标准差的计算结果见表 7.12.

表 7.12

长度(厘米)x_i	频数 f_i	f_ix_i(厘米)	$(x_i-\bar{x})$(厘米)	$(x_i-\bar{x})^2$(厘米²)	$f_i(x_i-\bar{x})^2$(厘米²)
1.2	2	2.4	−0.30	0.09	0.18
1.3	7	9.1	−0.20	0.04	0.28
1.4	10	14.0	−0.10	0.01	0.10
1.5	12	18.0	0.00	0.00	0.00
1.6	10	16.0	0.10	0.01	0.10
1.7	7	11.9	0.20	0.04	0.28
1.8	2	3.6	0.30	0.09	0.18
$\sum$	50	75.0 厘米			1.12 厘米²

$$\bar{x}=\frac{\sum f_ix_i}{n}=\frac{75.0\text{厘米}}{50}=1.50\text{厘米}$$

$$s=\sqrt{\frac{\sum f_i(x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{1.12\text{厘米}^2}{49}}=0.151186\text{厘米}\approx0.15\text{厘米}$$

7.16　利用(7.33)，(7.34)与(7.35)式，计算表 6.1 中长度样本的标准差.

解 ☞　若利用(7.33)，(7.34)与(7.35)式，需在表 6.1 中添加两列：x_i^2 和 $f_ix_i^2$. 修改后的表及标准差的计算结果见表 7.13.

表 7.13

长度(厘米)x_i	频数 f_i	f_ix_i(厘米)	x_i^2(厘米^2)	$f_ix_i^2$(厘米^2)
1.2	2	2.4	1.44	2.88
1.3	7	9.1	1.69	11.83
1.4	10	14.0	1.96	19.60
1.5	12	18.0	2.25	27.00
1.6	10	16.0	2.56	25.60
1.7	7	11.9	2.89	20.23
1.8	2	3.6	3.24	6.48
$\sum$	50	75.0 厘米		113.62 厘米^2

$$\bar{x}=\frac{\sum f_ix_i}{n}=\frac{75.0\text{ 厘米}}{50}=1.50\text{ 厘米}$$

利用(7.33)式

$$s=\sqrt{\frac{\sum f_ix_i^2-\frac{(\sum f_ix_i)^2}{n}}{n-1}}$$

$$=\sqrt{\frac{113.62\text{ 厘米}^2-\frac{(75.0\text{ 厘米})^2}{50}}{50-1}}=0.151186\text{ 厘米}\approx 0.15\text{ 厘米}$$

利用(7.34)式

$$s=\sqrt{\frac{n\sum f_ix_i^2-(\sum f_ix_i)^2}{n(n-1)}}$$

$$=\sqrt{\frac{(50\times 113.62\text{ 厘米}^2)-(75.0\text{ 厘米})^2}{50(50-1)}}=0.151186\text{ 厘米}\approx 0.15\text{ 厘米}$$

利用(7.35)式

$$s=\sqrt{\frac{\sum f_ix_i^2-n\bar{x}^2}{n-1}}$$

$$=\sqrt{\frac{113.62\text{ 厘米}^2-[50\times(1.50\text{ 厘米})^2]}{50-1}}=0.151186\text{ 厘米}\approx 0.15\text{ 厘米}$$

7.17 对于哪种类型样本，下述事实为真：$s^2=0$，$s^2=-1$，或$=-1$?

解 一个样本(或总体)的方差或标准差为零，当且仅当样本(或总体)的所有数据都相同. 样本(或总体)不会有负的方差或标准差，方差应永远为正值，因为它是基于离差平方的基础之上. 标准差应永远为正值，因为它是方差的正的平方根(见 7.9 节).

由分组频数分布计算近似标准差

7.18 对于表 6.10 中的第二次演讲比赛得分的分组频数分布，利用(7.37)式计算其近似样本标准差.

解 为利用(7.37)式，需在表 6.10 中添加两列：m_i^2 和 $f_im_i^2$，修改后的表及标准差的计算结果见表 7.14.

表 7.14

第二次比赛	组标 m_i	频数 f_i	f_im_i	m_i^2	$f_im_i^2$
45—49	47	1	47	2,209	2,209
50—54	52	0	0	2,704	0
55—59	57	4	228	3,249	12,996
60—64	62	3	186	3,844	11,532
65—69	67	5	335	4,489	22,445
70—74	72	3	216	5,184	15,552
75—79	77	6	462	5,929	35,574
80—84	82	11	902	6,724	73,964
85—89	87	9	783	7,569	68,121
90—94	92	17	1,564	8,464	143,888
95—99	97	5	485	9,409	47,045
$\sum$		64	5,208		433,326

$$\bar{x} \approx \frac{\sum f_im_i}{n} = \frac{5,208}{64} = 81.4$$

$$s \approx \sqrt{\frac{n\sum f_im_i^2 - (\sum f_im_i)^2}{n(n-1)}} = \sqrt{\frac{27,732,864 - 27,123,264}{64(63)}} = 12.2690 \approx 12.3$$

计算编码数据的方差和标准差

7.19 若 $c_i = a + bx_i$，证明 $s_x^2 = \frac{s_c^2}{b^2}$.

解 (6.16)式表明

$$\bar{x} = \frac{1}{b}(\bar{c} - a)$$

因而

$$\bar{c} = a + b\bar{x}$$

且

$$c_i - \bar{c} = (a + bx_i) - (a + b\bar{x}) = bx_i + a - a - b\bar{x} = b(x_i - \bar{x})$$

上式两边平方

$$(c_i - \bar{c})^2 = b^2(x_i - \bar{x})^2$$

两边求和，有

$$\sum(c_i - \bar{c})^2 = \sum b^2(x_i - \bar{x})^2 = b^2\sum(x_i - \bar{x})^2$$

上式两边同乘以 $\frac{1}{n-1}$

$$\frac{\sum(c_i - \bar{c})^2}{n-1} = \frac{b^2\sum(x_i - \bar{x})^2}{n-1}$$

于是有

$$s_c^2 = b^2 s_x^2$$

及

$$s_x^2 = \frac{s_c^2}{b^2}$$

7.20 考虑表 6.13 中的称重样本(单位:克)，利用(7.18)与(7.24)式直接计算 s_x^2 与 s_x. 然后，利用(7.39)，(7.40)及(7.41)式重新计算 s_x^2 与 s_x. 采用 $a=0$ 克，$b=0.0001$ 作为编码公式中的编码常量.

解 ☞　s_x^2 与 s_x 直接计算结果及利用编码公式与解码公式求得的结果见表 7.15.

表 7.15

重量(克)x_i	x_i^2(克²)	($c_i=0.0001x_i$)(克)	c_i^2(克²)
22,000.0	484,000,000	2.20	4.8400
24,000.0	576,000,000	2.40	5.7600
25,500.0	650,250,000	2.55	6.5025
27,500.0	756,250,000	2.75	7.5625
29,000.0	841,000,000	2.90	8.4100
30,000.0	900,000,000	3.00	9.0000
$\sum$ 158,000.0 克	4,207,500,000 克²	15.80 克	42.0750 克²

$$s_x^2=\frac{n\sum x_i^2-(\sum x_i)^2}{n(n-1)}=\frac{(6\times 4,207,500,000\text{ 克}^2)-(158,000.0\text{ 克})^2}{6(5)}=9,366,666.667\text{ 克}^2$$

$$s_x=\sqrt{s_x^2}=\sqrt{9,366,666.67\text{ 克}^2}=3,060.5010\text{ 克}\approx 3,060.50\text{ 克}$$

利用编码和解码公式计算

$$s_c^2=\frac{n\sum c_i^2-(\sum c_i)^2}{n(n-1)}=\frac{(6\times 42.0750\text{ 克}^2)-(15.80\text{ 克})^2}{6(5)}=0.0936667\text{ 克}^2$$

$$s_c=\sqrt{s_c^2}=\sqrt{0.0936667\text{ 克}^2}=0.306050\text{ 克}$$

对 s_c^2 解码：$s_x^2=\frac{s_c^2}{b^2}=\frac{0.0936667\text{ 克}^2}{(0.0001)^2}=9,366,670.000\text{ 克}^2$

对 s_c 解码：$s_x=\frac{s_c}{b}=\frac{0.306050\text{ 克}}{0.0001}=3,060.5000\text{ 克}\approx 3,060.50\text{ 克}$

7.21　习题 6.9 中讨论的两个工厂 A，B，开始时有相同的平均每小时薪金($\bar{x}=\$5.19$)，后来，采用不同的提薪方式，即工厂 A 的雇员每小时薪金增加 5%，从而平均每小时薪金增加到 \$5.66；而工厂 B 的雇员每小时薪金增加 \$0.05，从而平均每小时薪金增加到 \$5.44，非常巧合的是，两个工厂，在增加薪金之前，对他们的每小时薪金，有相同的样本标准差($s_x=\$0.40$)．采用下述公式求两个工厂在增加薪金后的标准差：$s_c=bs_x$．增加薪金后，哪个工厂有较大的偏离度？

解 ☞　对于工厂 A，若 x_i 表示以前的每小时薪金，则每个雇员现在的每小时薪金为：$c_i=\$0.00+1.05x_i$．记 s_x 表示加薪前的标准差，s_c 表示加薪后的标准差，编码常量为 $a=\$0.00$，$b=1.05$，则

$$s_c=b\,s_x=1.05(\$0.40)=\$0.42$$

对于工厂 B，每位雇员现在的每小时薪金为：$c_i=\$0.05+x_i$，编码常量为 $a=\$0.05$，$b=1.00$．因而

$$s_c=b\,s_x=1.00(\$0.40)=\$0.40$$

于是，对于工厂 A，其标准差由 \$0.40 增加到 \$0.42，对于工厂 B，其标准差保持不变(\$0.40)．因而，加薪后工厂 A 比工厂 B 有较大的偏离度．

Chebyshev 定理

7.22　下述描述性度量来自于一个时间测量的样本：$n=400$，$\bar{x}=21.2$ 秒，$s=1.7$ 秒．利用 7.15 节中的 Chebyshev 定理回答下述问题：这组数据中，距离均值 $2\frac{3}{4}$ 个标准差的范围之内所占的比例至少为多少？

解 ☞　位于区间"均值 $\pm k$ 个标准差"内(这里 $k=2\frac{3}{4}$)的数据的比例至少为

$$1-\frac{1}{k^2}=1-\frac{1}{\left(2\frac{3}{4}\right)^2}=1-\frac{16}{121}=0.867769$$

7.23 下述度量是一个长度测量的样本:$n=10000$,$\bar{x}=20.0$ 厘米,$s^2=0.25$ 厘米^2. 利用 Chebyshev 定理,求该样本中落入区间 19.0 厘米至 21.0 厘米内的观测值至少有多少个?

解 ☞ 先求 s

$$s=\sqrt{s^2}=\sqrt{0.25\text{ 厘米}^2}=0.5\text{ 厘米}$$

为求 k,已知区间(19.0 厘米至 21 厘米)为

$$(\bar{x}=20.0\text{ 厘米})\pm(ks=1.0\text{ 厘米})$$

因此

$$k=\frac{1.0\text{ 厘米}}{s}=\frac{1.0\text{ 厘米}}{0.5\text{ 厘米}}=2$$

由 Chebyshev 定理知,位于区间"$\bar{x}\pm2s$"内的数据的比例至少为

$$1-\frac{1}{k^2}=1-\frac{1}{2^2}=0.75$$

因而,该样本落入区间 19.0 厘米至 21.0 厘米内至少有

$$0.75(n=10,000)=7,500\text{ 个观测值}$$

7.24 某工厂制造一种手电筒电池,该工厂声称其制造的电池中,至少有 96%的电池寿命为 95 小时至 105 小时,该厂随机抽取 1000 个电池,测得平均寿命为 $\bar{x}=100$ 小时. 若该厂的声明是正确的,求样本标准差 s 的最大可能值?

解 ☞ 由 Chebyshev 定理知,若数据的 96%以上位于区间 $\bar{x}\pm ks$ 内,则

$$1-\frac{1}{k^2}=0.96$$

$$\frac{1}{k^2}=1-0.96=0.04$$

$$k=\frac{1}{0.2}=5$$

如果区间(95 小时到 105 小时)[即($\bar{x}=100$ 小时)±5 小时]至少包括数据的 96%,则

$$ks=5\text{ 小时}$$

$$s=\frac{5}{k}=\frac{5}{5}=1\text{ 小时}$$

因而 $s=1$ 小时是样本标准差的最大可能值.

7.25 某食品公司出售一种每袋 12 盎司(1 盎司=28.35 克)的薯条,按规定,每袋必须至少装 12 盎司,他们设计了一种自动包装机,每袋装 12.20 盎司. 为检验该包装机,随机抽取 500 袋,测得结果为:$n=500$,$\bar{x}=12.20$ 盎司,$s=0.04$ 盎司. 若想使产品中至少 99%符合规定,应对机器作何调整?

解 ☞ 为利用 Chebyshev 定理,将此问题转化为:这些数据中至少有 99%位于区间 $12.20\pm k(0.04)$盎司时,k 为多少?

由定理知

$$1-\frac{1}{k^2}=0.99$$

$$k=10$$

然而当 $k=10$ 时,该区间将延伸到 12 盎司以下(12.20 盎司$\pm10(0.04)$盎司),小于所必需的 12 盎司. 故,若想保证至少 99%的产品符合规定,该公司必需调整机器或者增加每袋的平均重量($\bar{x}$),或者减少每袋重量的离散程度(s).

经验法则

7.26 假定总体服从正态分布 $N(\mu,\sigma^2)$,求总体中小于 $\mu-2\sigma$ 的百分比?

解 ☞ 利用例 7.15 中的推导,小于 $\mu-2\sigma$ 的百分比为

$$\frac{1}{2}(100\%-95.4\%=4.6\%)=2.3\%$$

7.27 某生物学家正在研究一种蜗牛，他随机抽取一个容量为500的样本，测量它们外壳的直径(单位：毫米)，假定直径基本上服从正态分布，利用经验法则(见7.16节)，确定外壳直径位于区间 $\bar{x}\pm s$ 内的蜗牛的近似个数.

解☞ 由7.16节知，大约有68%的数据位于区间 $\bar{x}\pm s$ 内，因此，外壳直径位于该区间内的蜗牛个数大约是 $0.68\times500=340$.

7.28 为什么有的统计书中将数据集的异常值定义为距离其均值三个标准差的一个测量值?

解☞ 有时在一个数据集中存在一些远远大于或者远远小于其他数据的值，这种极端值称为异常值，其一般定义为：与均值的差超过三个标准差的数据. 由经验法则的分布百分比，可以看出这样定义的理由. 经验法则指出，不论是对称分布，还是有偏斜的分布，大约100%的数据位于距离均值三个标准差的范围之内. 因而，某个数据位于三个标准差之外认为是特别不寻常的事件——一个异常值，这种异常值可以是由于操作不当(观测，记录，计算)，或者试验失败，或者是某个复杂的外部变量，如来自另一总体的测量值. 如果存在异常值，而且可以认为它是由于某些方法问题而引起的，有时可从数据集中剔除这个异常值.

7.29 假定某数据集近似服从正态分布，问其极差与其标准差之间的关系?

解☞ 对于此数据集，经验法则指出，大约95%的数据位于区间(均值±2标准差)之内，大约100%的数据位于区间(均值±3标准差)之内. 因而，作为一种经验，可以假定该数据集的极差约在4至6个标准差之间，即标准差应处于极差的 $\frac{1}{4}$ 至 $\frac{1}{6}$ 倍之间. 这种关系可以用来迅速地检查标准差计算的准确性.

集中趋势和偏离性的图示

7.30 某治疗失眠诊所的医生正在研究一种新安眠药，20位患者志愿参加试验，该医生利用随机数表(见3.23节)将他们随机分成两组：药物组与控制组，每组10人，记录下所有20位患者睡在诊所里一个晚上的脑电波. 药物组的每个患者服用这种新药加一杯牛奶，而控制组的每位患者服用一种糖剂(安慰剂)加一杯牛奶，然后根据其脑电波观测他们需多长时间才能入睡，即从晚上10点(0分钟)到脑电波中的第一个入睡信号出现止. 两组记录结果如图7-6所示，其中 Y 轴表测量值单位，每个条形高度表示组平均 $\bar{x}$，条形上方的垂直线段长度表示标准差 s. 利用图中所给的信息，回答下述问题：(a)根据Chebyshev定理，求至少包含每组数据68%的时间区间? (b)根据经验法则，求大约包含每组数据68%的时间区间?

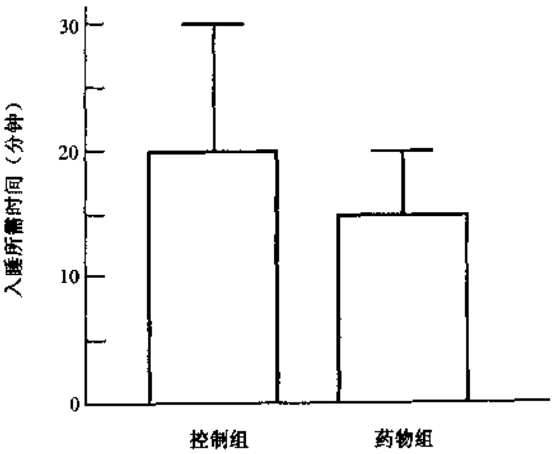


图7-6

解☞ (a)由Chebyshev定理知，若至少有68%的数据位于区间 $\bar{x}\pm ks$ 内，则

$$1-\frac{1}{k^2}=0.68$$

$$k=1.76777$$

对于每一组，$\bar{x}$ 与 s 可通过对条形图目测或测量而得到，结果为：对于控制组，$\bar{x}=20$ 分钟，$s=10$ 分钟；对于药物组，$\bar{x}=15$ 分钟，$s=5$ 分钟. 因而对于控制组，包含至少68%的数据的区间为

$$\bar{x}\pm ks = 20\text{ 分钟}\pm(1.76777\times10\text{ 分钟})$$

$$=20\text{ 分钟}\pm17.6777\text{ 分钟}$$

$$\approx (2.3\text{分钟}, 37.7\text{分钟})$$

对于药物组,该区间为

$$\bar{x} \pm ks = 15\text{分钟} \pm (1.76777 \times 5\text{分钟})$$
$$= 15\text{分钟} \pm 8.83885\text{分钟}$$
$$\approx (6.23\text{分钟}, 23.8\text{分钟})$$

(b)假定每组的时间近似服从正态分布.经验法则表明大约68%的数据位于区间 $\bar{x} \pm s$ 内.对于控制组

$$\bar{x} \pm s = 20\text{分钟} \pm 10\text{分钟} \approx (10\text{分钟}, 30\text{分钟})$$

对于药物组

$$\bar{x} \pm s = 15\text{分钟} \pm 5\text{分钟} \approx (10\text{分钟}, 20\text{分钟})$$

注:由这些结果看来这种安眠药是有效的,因为平均地说,药物组的患者睡眠更快,所需时间更集中.然而,这些结果仅刻画了小样本情形,若想得到更一般的结论,我们还需对所有可能服药与未服药的人做试验,若想在某种确信(概率)水平基础上做出这种比较,则需用到下一章介绍的一些统计推断的方法.

7.31 利用图 7-4(b)的漂浮矩形图形式将图 7-6 中的数据表示成图 7-7,根据图 7-7 回答下述问题:(a)抽样分布是否对称?(b)每个样本是否包含异常值?

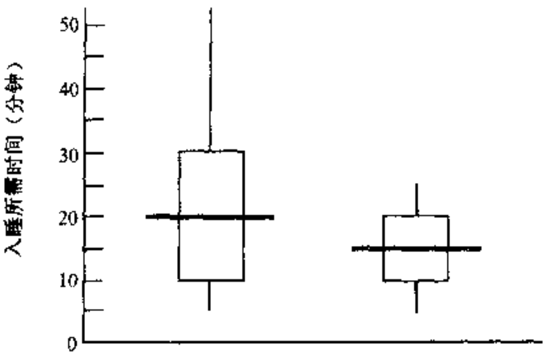


图 7-7

解 (a)药物组的时间分布关于均值对称,而控制组的时间分布向正方偏斜(沿着纵轴)(b)在习题 7.28 中,我们给出了异常值的一般定义:距离均值超过三个标准差的数据.这里,对于药物组不存在异常值,而对于控制组,至少有一个异常值:$x_i = 55$ 分钟,它距离均值 $3\frac{1}{2}$ 标准差.

注:尽管图 7-6 的条形图是一种常用表示法,但若想强调非对称性或极端值,图 7-7 的方法更有用,这里就是一例.正如习题 7.30(b)所指出的那样,控制组的时间分布不是如假定的正态分布,因而经验法则不适用.

7.32 由于安眠药试验结果(见图 7-6 及 7-7)表明这种新药对于减少入睡所需时间很有效,为更好预测,我们想研究连续四天的药效,20 位患者被随机分成新药物组与控制组(每组 10 人),所有患者在连续四天内重复(习题 7.30 描述的)试验,药物组的试验结果如图 7-8 所示,这是类似于图 7-4(c)的线形图.利用 Chebyshev 定理,对于第一天和第四天分别分别确定至少包含 68%数据的时间区间?

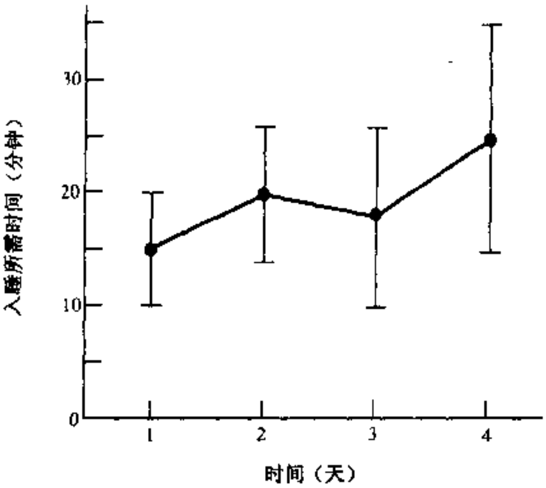


图 7-8

解 第一天,$\bar{x} = 15$ 分钟,$s = 5$ 分钟,且由习题 7.30(a)知,包含至少 68%数据的区间的 $k = 1.76777$.因而,包含至少 68%数据的区间为

$$\bar{x} \pm ks = 15\text{分钟} \pm (1.76777 \times 5\text{分钟})$$
$$= 15\text{分钟} \pm 8.83885\text{分钟}$$

$$\approx (6.2 \text{ 分钟}, 23.8 \text{ 分钟})$$

第四天，$\bar{x}=25$ 分钟，$s=10$ 分钟，故所求区间为

$$\begin{aligned} \bar{x} \pm ks &= 25 \text{ 分钟} \pm (1.76777 \times 10 \text{ 分钟}) \\ &= 25 \text{ 分钟} \pm 17.76777 \text{ 分钟} \\ &\approx (7.3 \text{ 分钟}, 42.7 \text{ 分钟}) \end{aligned}$$

7.33 习题 7.32 中药物组的试验结果似乎表明：在连续四天内，这种新药的效果有所降低，因为 $\bar{x}$ 与 s 都增大了．为将此结果与控制组进行比较，将控制组也表示在图 7-8 中，用虚线连结的空心点表示，见图 7-9．利用 Chebyshev 定理，对于控制组的第一天和第四天确定至少包含 68% 数据的时间区间．

解 ☞　在第一天及第四天，$\bar{x}=20$ 分钟，$s=10$ 分钟，为对于包含至少 68% 的数据的区间，$k=1.76777$．因而

$$\begin{aligned} \bar{x} \pm ks &= 20 \text{ 分钟} \pm (1.76777 \times 10 \text{ 分钟}) \\ &= 20 \text{ 分钟} \pm 17.76777 \text{ 分钟} \\ &\approx (2.3 \text{ 分钟}, 37.7 \text{ 分钟}) \end{aligned}$$

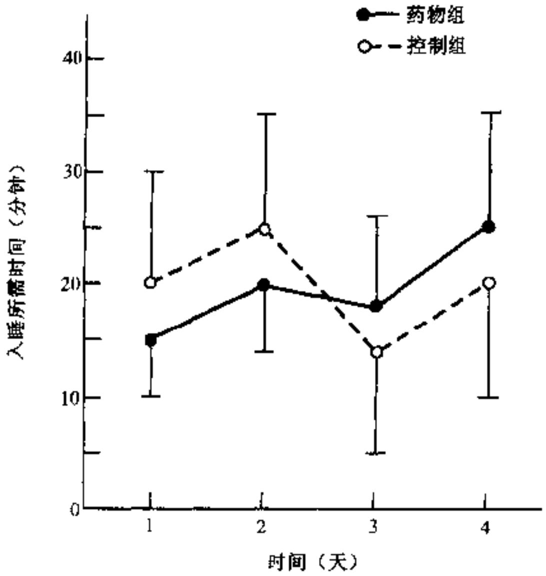


图 7-9

注：由图 7-9，可以在描述性统计水平上比较两组的试验结果，我们看到控制组没有特殊的模型(在四天内以稳定的离散性 $s=10$ 分钟上下波动)，而药物组则表现出 $\bar{x}$ 与 s 有稳步增长的趋势，至第四天，达到较高的 $\bar{x}$ 且与控制组有相同的 s，这些结果表明药物组已适应了新药：服用同样剂量该药，随时间增长而使得药效降低．然而，若想得出这样一般的结论，必需用推断统计方法分析这些分组试验结果．

变异系数

7.34 某石油公司随机抽取 1,000 名顾客作调查，就某一周的情况询问了如下的问题：你所驾驶的里数？你所购买的汽油量(单位：加仑)？你购买汽油的花费？调查结果如下：$\bar{x}=120$ 英里，$s=9.6$ 英里；$\bar{x}=6$ 加仑，$s=0.5$ 加仑；$\bar{x}=\$6.30$，$s=\$0.70$．利用(7.45)式计算这些观测值的相对离散性．

解 ☞　对于每周的英里数

$$V=\frac{s}{\bar{x}}(100\%)=\frac{9.6 \text{ 英里}}{120 \text{ 英里}}(100\%)=8.0\%$$

对于每周的加仑数

$$V=\frac{s}{\bar{x}}(100\%)=\frac{0.5 \text{ 加仑}}{6 \text{ 加仑}}(100\%)=8.3\%$$

对于每周的花费

$$V=\frac{s}{\bar{x}}(100\%)=\frac{\$0.70}{\$6.30}(100\%)=11.1\%$$

由此可以看出变异系数的一个重要应用：由于它不依赖于测量单位，变异系数可以用来比较具有不同单位的分布的相对离散程度．这里我们可以看到，每周的花费与其他两个量相比，具有较大的相对离散性．

7.35 用°F 或℃测量的温度的变异系数，为什么不恰当？

解 ☞　由于变异系数是一个比值，它只适用于比率水平测量(见 2.7 节)，而温度(单位：°F或：℃)是间隔水平测量．

标准分和标准化变量

7.36 某大学体育系安排你负责该系学生的学习情况,一位足球运动员告诉了他的学习成绩(百分制):在化学期中考试中,他考了 68 分,而全班的成绩为:$\bar{x}=74, s=13$;在历史期中考试中,他考了 74 分,而全班成绩为:$\bar{x}=84, s=7$. 问相对于全班的平均成绩,他哪门课更好?

解 ☞　对于两门期中考试,分别计算标准分

化学:$z_i=\dfrac{x_i-\bar{x}}{s}=\dfrac{68-74}{13}=-0.461538\approx-0.46$;

历史:$z_i=\dfrac{x_i-\bar{x}}{s}=\dfrac{74-84}{7}=-1.42857\approx-1.43$.

由标准分,显然该学生在化学期中考试中相对较好,因为化学期中考试,他的成绩比平均成绩低 0.46 个标准差,而历史期中考试,他的成绩比平均成绩低 1.43 个标准差.

7.37 任意给定一个来自于标准化变量 Z 的样本,证明:(a) $\bar{z}=0$,(b) $s_z=1$.

证明:(a)由于

$$z_i=\frac{x_i-\bar{x}}{s}$$

故

$$\bar{z}=z_i\text{ 的均值}=\frac{\sum z_i}{n}=\frac{\sum\left(\frac{x_i-\bar{x}}{s}\right)}{n}=\frac{1}{ns}\sum(x_i-\bar{x})$$

因而,由

$$\sum(x_i-\bar{x})=0$$

得

$$\bar{z}=\frac{1}{ns}(0)=0$$

(b) $s_z=z_i$ 的标准差 $=\sqrt{\dfrac{\sum(z_i-\bar{z})^2}{n-1}}$

因而

$$s_z=\sqrt{\frac{\sum(z_i-0)^2}{n-1}}=\sqrt{\frac{\sum z_i^2}{n-1}}=\sqrt{\frac{\sum\left(\frac{x_i-\bar{x}}{n-1}\right)^2}{n-1}}$$

$$=\sqrt{\left(\frac{1}{s^2}\right)\left(\frac{1}{n-1}\right)\sum(x_i-\bar{x})^2}$$

$$=\sqrt{\left(\frac{1}{s^2}\right)(s^2)}$$

$$=1$$

注:类似地可以证明对于任意标准化变量的总体,有 $\mu_z=0, \sigma_z=1$. 标准化变量的均值和标准差的这些性质很重要,在第 12 章中我们将研究服从正态分布的标准化变量,称之为**标准正态分布**(或**标准化正态分布**).

四分位极差和四分位差

7.38 对于表 6.7 中的长度数据,已知 $Q_2=\tilde{x}=1.50$ 厘米(见例 6.13),四分位数中点=1.50 厘米(见例 6.15),四分位差=0.12 厘米(见例 7.19). 求位于下列区间内的数据所占的百分比:(a) (四分位数中点)±(四分位差);(b) $\tilde{x}$±(四分位差)?

解 ☞　由于这是一个单峰对称分布,Q_1,Q_3 的中点既等于四分位数中点,又等于 $\tilde{x}$,因而(a)、(b)定义的区间包含了所给数据的中间的 50%.

7.39 考虑表 6.14 中的重量数据,已知 $Q_1=17.3056$ 克,$Q_2=\tilde{x}=18.2917$ 克,$Q_3=19.0$ 克,四分位数中点=18.1528 克(见习题 6.20 和 6.25). 试计算四分位极差和四分位差,并

确定位于下述区间内的数据所占百分数:(a)(四分位数中点)±(四分位差);(b)$\tilde{x}$±(四分位差).

解

$$四分位极差 = Q_3 - Q_1 = 19.0\ g(克) - 17.3056\ g = 1.6944\ g \approx 1.7\ g$$

$$四分位差 = \frac{Q_3 - Q_1}{2} = \frac{1.6499\ g}{2} = 0.8472\ g \approx 0.85\ g$$

(a)对于任意形状的分布,从Q_1到Q_3的区间的中点均等于四分位数中点,因而,对于这个单峰向左偏斜的分布,区间(四分位数中点)±(四分位差)包含该分布的中间的50%.

(b)对于这种非对称分布,已知$\tilde{x}\neq$(四分位数中点).因而,仅采用这些简单方法无法确定位于区间$\tilde{x}$±(四分位差)内的数据所占百分比.

注:由于对于任意形状分布,四分位差总为包含数据的中间50%的区间的一半的一种度量,故在描述严重偏斜或多峰分布的离散性时,常被用来代替标准差,此时中位数或四分位数中点常代替均值,作为位置的一种度量.

7.40 对于表6.15给出的温度的数据,已知$Q_1 = 100.8333°F$,$Q_2 = \tilde{x} = 101.3889°F$,$Q_3 = 102.3°F$,四分位数中点$= 101.5667°F$(见习题6.21,6.26及6.51),试计算四分位极差和四分位差,并确定表6.15的数据中位于下列区间内的百分比:

解

$$四分位极差 = Q_3 - Q_1 = 102.3°F - 100.8333°F = 1.4667°F \approx 1.5°F$$

$$四分位差 = \frac{Q_3 - Q_1}{2} = \frac{1.4667°F}{2} = 0.7333°F \approx 0.73°F$$

(a) 对于这种单峰正向偏斜的分布,区间(四分位数中点)±(四分位差)包括了分布的中间50%的数据.

(b) 对于这种有偏斜的分布,已经证明$\tilde{x}\neq$(四分位数中点).因而我们不能用这些方法确定位于区间$\tilde{x}$±(四分位差)内的数据所占的百分比.

盒子图与五数概括

7.41 利用习题6.20所给信息,对于表6.8中的重量数据,构造一个盒子图并给出五数概括.

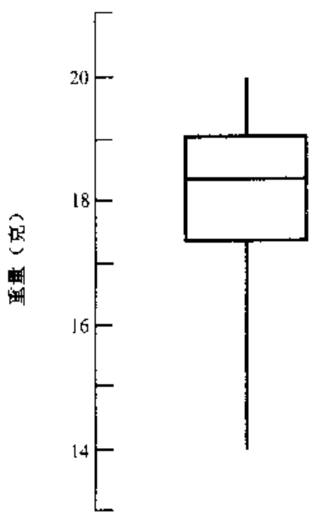


图7-10

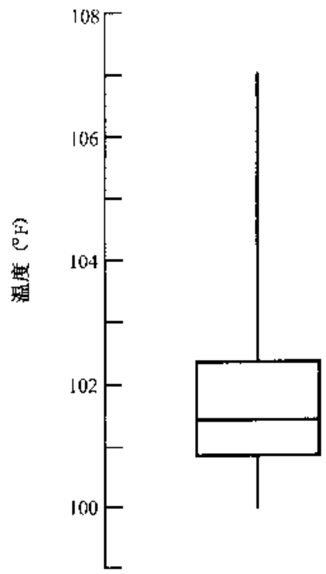


图7-11

解 表示五数Q_1、Q_2、Q_3,x_1、x_n的盒子图见图7-10.

7.42 利用习题6.21和6.51所给信息,对于表6.9给出的温度数据,构造一个盒子图,表示五数概括.

解 ☞　表示五数 Q_1, Q_2, Q_3, x_s, x_l 的盒子图见图 7-11.

7.43　利用习题 6.23 和 6.53 给出的精确四分位数及表 6.11 所给信息,构造一个盒子图,表示五数概括.

解 ☞　表示五数 Q_1, Q_2, Q_3, x_s, x_l 的盒子图见图 7-12.

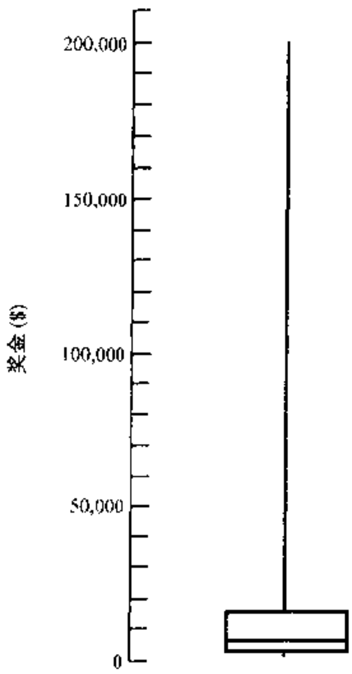


图 7-12

补充习题

平均偏差

7.44　计算样本:0.0, 15.3, 0.0, 100.3, 6.21, 0.0 的极差与平均偏差.

答案:极差 = 100.3,平均偏差 = 34.3889≈ 34.39

7.45　对于表 6.9 给出的样本,计算其极差,再利用(7.6)式,计算其平均偏差.

答案:极差 = 7°F,平均偏差 = 1.0470°F≈1.0°F

总体方差

7.46　利用(7.12)及(7.14)式计算总体:0, 1, 2, 11, 19 的方差.

答案:53.84

7.47　利用(7.12)及(7.15)式计算总体:0.00, 0.01, 0.02, 0.11, 0.19 的方差.

答案:0.005384≈ 0.0054

样本方差

7.48　利用(7.16)及(7.17)式计算样本:138, 129, 132 的方差.

答案:21.00

7.49　利用(7.16)及(7.19)式计算样本:152, 129, 148 的方差.

答案:151.00

总体标准差

7.50　利用(7.20)及(7.22)式计算习题 7.46 所给总体的标准差.

答案:7.3

7.51 利用(7.20)及(7.23)式计算习题 7.47 所给总体的标准差.

答案:0.073

7.52 对表 6.11 中的高尔夫球比赛奖金的数据,利用(7.30)式计算该总体的标准差.

答案:\$28,536.14≈\$28,500

样本标准差

7.53 利用(7.27)计算习题 6.37(a)—(d)中所给样本的标准差.

答案:(a) 0 秒,(b)4.47214 秒≈4.5 秒,(c)40.9383 公里,(d)5.2743 公里

7.54 对于表 6.8 中的重量数据,利用(7.34)式计算该样本的标准差.

答案:1.24847 克≈1.2 克

7.55 对于 6.9 中的温度数据,利用(7.35)式计算该样本的标准差.

答案:1.4485°F≈1.4°F

由分组频数分布计算近似标准差

7.56 对于表 6.12 中高尔夫球比赛奖金的分组频数分布,利用(7.36)式计算其近似总体标准差.

答案:$\sigma\approx$ \$28,214.58≈\$28,200

计算编码数据的方差和标准差

7.57 对于下述样本:0.0013, 0.0027, 0.0039, 0.0022,首先直接利用(7.27)式计算其标准差,然后利用(6.14),(7.39)和(7.41)式重新计算其标准差.在编码公式中,采用 $a=0$, $b=1000$ 作为编码常量.

答案:0.0010844≈0.00108

Chebyshev 定理

7.58 对于习题 7.22 给出的样本,利用 Chebyshev 定理(见 7.15 节),回答如下问题:至少有百分之多少的数据位于区间(19.3 秒,23.1 秒)内?

答案:至少 19.9%的数据位于该区间内

7.59 对于习题 7.22 给出的样本,利用 Chebyshev 定理,回答如下问题:至少有百分之多少的数据位于区间($\bar{x}=21.2$ 秒)±4.1 秒内?

答案:至少 82.8%的数据位于该区间内

经验法则

7.60 对于习题 7.27 外壳直径的样本,利用经验法则(见 7.16 节),确定位于区间 $\bar{x}\pm 2s$ 内的数据的近似个数.

答案:大约 0.95×500=47 个

7.61 对于习题 7.27 外壳直径的样本,利用经验法则(见 7.16 节),确定位于区间 $\bar{x}\pm 3s$ 内的数据的近似个数.

答案:大约 500 的 100%个

7.62 若总体服从正态分布,则总体的百分之多少位于区间 $\mu\pm 2\sigma$ 内?

答案:97.7%

集中趋势和偏离性的图示

7.63 图 7-7 中,哪个样本有最小值和最大值?

答案:两个样本的 x_s 均为 5 分钟,而控制组的 x_l 较大(55 分钟对 25 分钟)

7.64 在习题 7.29 中,我们指出,作为一种经验法则,标准差应在$\frac{1}{4}$(极差)与$\frac{1}{6}$(极差)之间.对于图中的控制组与药物组这一法则是否正确?

答案:控制组:$s\approx\frac{1}{5}$(极差);药物组:$s\approx\frac{1}{4}$(极差)

变异系数

7.65　在过去的一年里(共250个交易日),公司A的股票,每股平均日交易价为:$\mu=\$140$,$\sigma=\$8$;而公司B的股票,每股平均日交易价为$\mu=\$5$,$\sigma=\$0.8$. 利用(7.43)式确定哪种股票有较大的相对偏离程度?

答案:公司A:$V=5.7\%$;公司B:$V=16.0\%$

标准分与标准化变量

7.66　假定样本$x_1,x_2,\cdots,x_n$近似服从正态分布,(a)利用Chebyshev定理(见7.15节)确定样本中标准值在-2与2之间的数据至少有多少?及(b)利用经验法则(见7.16节)确定样本中标准值在-1与1之间的数据至少有多少?

答案:(a) 0.75,(b)$\approx 68\%$

四分位极差与四分位差

7.67　考虑表6.7、6.14及6.15中的分布,数据的百分之多少位于区间(四分位数中点)±(四分位差)的外部(之上和之下)?

答案:50%

盒子图与五数概括

7.68　图7-13给出的盒子图表示了一种分布,该分布是由五数概括:Q_1,Q_2,Q_3,x_s,x_l表示的. 试由该盒子图确定:四分位极差,四分位差,$\tilde{x}$,极差,该分布是否对称?

答案:四分位极差=4秒,四分位差=2秒,$\tilde{x}=2$秒,极差=10秒,分布是对称的

7.69　图7-14给出的盒子图表示了一种分布,该分布是由五数概括:Q_1,Q_2,Q_3,x_s,x_l表示的. 试由该盒子图确定:四分位极差,四分位差,$\tilde{x}$,极差,该分布是否对称?

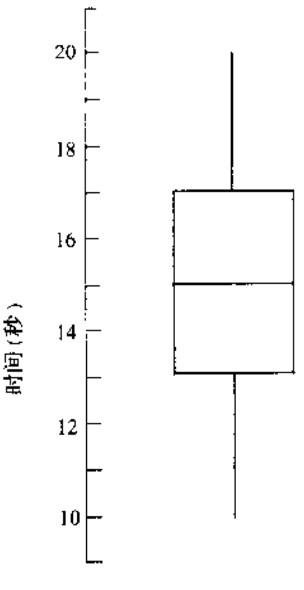


图7-13

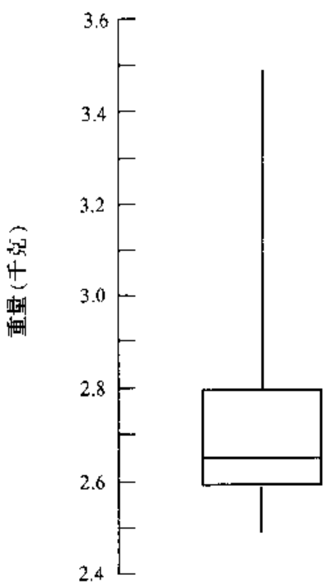


图7-14

答案:四分位极差=0.2千克,四分位差=0.1千克,$\tilde{x}=2.65$千克,极差=1.0千克,分布是正向偏斜的

第八章　概率：古典解释，相对频数解释，集合论解释和主观解释

8.1　概率的古典解释

概率作为统计科学概述性介绍的一部分，在第3章(见3.16节)已讨论过. 概率是推断性统计学的一个基本组成部分，用来计量从样本对总体进行推断(见3.5节)的不确定性的程度. 因此，随着本书的重点从描述性统计学转移到推断性统计学，有必要在本章和下一章首先讲述概率的基本概念和基本方法.

在3.16节指出，事件的概率(0和1之间的数)表明事件发生的可能性或机会，其中，0表示事件不可能发生，1表示事件必然发生. 然而，根据不同计算形式，概率有四种不同解释：**古典解释**(本节)、**相对频数解释**(见8.2节)，**集合论解释**(见8.3至8.6节)和**主观解释**(见8.7节).

我们使用标准术语系统讨论概率的上述4种解释，用到以下概念：**试验，一次试验，结果**和**事件**. 科学研究中的一个试验是指通过控制一个独立变量来了解它对另一个相依变量(见1.19节)的影响. 然而，在统计学中，试验是指产生测量值的任意过程. 例如，掷一粒六面的骰子并在骰子停止转动后观测朝上的点数，就是一个统计学试验. 一个试验的每一次相同重复称为该试验的一次试验(或**子试验**)；一次试验的每一个不同的最终表现，即产生的每一个不同测量值，称为一个结果. 于是，对六面骰子每掷一次就是掷骰子试验的一次试验，并且每次试验产生6个可能结果(1,2,3,4,5,6点)之一，这些结果表示离散比例测量水平的若干种类(见2.8节). 一次试验的任何确定结果都能归为不同的类，这些不同的类称为事件. 例如，在掷骰子试验中，掷出某个确定点数(如3点)即是该试验的一个可能事件，同样，掷出某个偶数点(2,4,或6点)也是一个事件.

概率的古典解释在4种解释中最为古老，早在17世纪已经提出对赌博中机会的研究对策. 古典概率处理的是**理想化对策**：在对策中，每一次试验都是在一致且理想的条件下进行. 这样的理想化对策总是"公平的"，因为所有的可能结果都是**等可能地出现**. 诸如掷骰子试验，如果骰子密度均匀，表面上的点洞完全对称，并且在每次试验中用相同的手势和力度掷出，就是上面所言的一个理想化对策. 在这样的条件下，骰子停止转动后，六面有相等的朝上的可能性. 结果的等可能性是古典解释的显著特征.

在理想化的试验中，如果试验的结果**有利于** A，则称事件 A 发生. 于是，对于掷骰子试验，设 A 是掷出3点，如果有利于这一事件(掷出3点)的某个结果发生，则 A 发生. 类似地，设 A 是掷出偶数点，如果3个有利结果(2,4,或6点)的任一个发生，则 A 发生. 对这类试验，**事件 A 的概率**[用 $P(A)$ 表示]是有利于 A 的可能结果个数(用 N_A 表示)与试验的可能结果总数(用 N 表示)的比，即

$$P(A)=\frac{N_A}{N} \tag{8.1}$$

这是公式(3.3)的符号表示，这里假设所有的可能结果等可能地出现.

对于掷骰子试验，如果 A 是掷出3点，则在总共6个可能结果中仅有1个有利结果，于是

$$P(A)=\frac{N_A}{N}=\frac{1}{6}=0.17,\quad 或\ 17\%$$

如果 A 是掷出偶数点，则在总共6个可能结果中有3个有利结果，于是

$$P(A)=\frac{N_A}{N}=\frac{3}{6}=0.50,\quad 或50\%$$

以上计算表明，概率可以写成简单的分数，小数，或百分数. 如果将分数变换为小数或百分数，则概率通常以两个有效数字表示.

古典解释适用于任何这样的试验：事先已知试验的所有可能结果；对每次试验，必有一个结果发生，并且所有结果等可能地出现. 在古典解释中，因为概率在试验进行之前即已明确，所以古典概率也称为**先验**(事前)**概率**.

例 8.1　有一枚均匀，完全平坦且对称的硬币，并有两面：正面和反面. 试验是将该硬币抛入空中并观测着地后哪面朝上. 计算以下概率：(a)P(正面)，(b)P(反面)，(c)P(正面或反面).

解 ☞　(a)对于这个试验，每一次试验有两个等可能的结果：正面或反面，因此 $N=2$. 这里 $N_A=1$，因为两个结果中仅有一个(得到正面)有利于感兴趣的事件. 因此，用公式(8.1)

$$P(正面)=\frac{N_A}{N}=\frac{1}{2}=0.50,\quad 或50\%$$

(b)解答与(a)相同

$$P(反面)=\frac{N_A}{N}=\frac{1}{2}=0.50,\quad 或50\%$$

(c)这里，N 同上；但 $N_A=2$，因为有两个可能结果有利于感兴趣的事件

$$P(正面或反面)=\frac{N_A}{N}=\frac{2}{2}=1.0,\quad 或100\%$$

这个解答表明，对一枚硬币进行理想化投掷，出现或者是正面或者是反面的概率是 100%.

8.2　概率的相对频数解释

概率的古典解释存在局限性：仅适用于理想化试验，即所有可能结果事先已知并且等可能地出现. 然而，有许多实例不满足古典解释的上述要求，也需要确定其中事件的发生概率. 在以下例子中，尽管每个试验都只有两个可能结果(是或否)，但是不能假设两个结果等可能地出现：(1)保险公司的统计学家，计算年龄为 45 岁的男子能够活到 46 岁的概率；(2)地质学家计算位于地质断层的某城市在随后 5 年内将发生一次地震的概率；(3)汽车前灯制造商计算 100 个前灯的一批中没有不合格品的概率. 对于这些实例，尽管古典概率无法处理所要求的概率，但是可以使用**相对频数解释**(也称为**频率解释，或经验解释**)进行计算.

在古典解释中，概率在试验进行之前就已明确. 在相对频数解释中，概率是根据以前试验产生的结果进行计算. 对多次重复相同或"相似"试验得到的数据进行分析，可以获得感兴趣事件发生的**相对频数**(即次数的比). 从以前数据获得的事件发生的相对频数，在之后的试验中可以看作是对将来该事件发生概率的一个估计. 于是，在上述的保险实例中，45 岁的男子在 46 岁之前死亡的相对频数往往从历史数据计算得到. 类似地，对于地震实例，地质学家往往首先找出与问题所述城市相似的情形，然后计算 5 年内这种情形下发生地震的相对频数. 同样，汽车前灯制造商往往要测试若干个 100 个前灯的样本，并计算样本中不合格品的相对频数.

为给出相对频数解释的基本概率公式，考虑下面的简单例子. 你认为某枚硬币是"不公正"的——可能存在某种控制方法，使得掷一次硬币的两个可能结果(正面或反面)不是等可能地出现. 为研究这种可能性，你掷了该硬币 100 次，得到 70 次正面和 30 次反面. 出现正面的相对频数是 $70/100=7/10=0.70$，或 70%，于是 70%就是你对将来投掷该硬币一次得到正面的概率的估计. 这只是一个"估计"，即一个近似值，因为：如果继续掷硬币达到 1000 次或 1000000 次，你很有可能得到一个不同结果.

概率的相对频数解释可以表述为：

事件 A 的概率[仍用 $P(A)$ 表示]近似等于在一系列试验中 A 发生的次数(用 n_A 表示)对试验总次数(用 n 表示)的比. 即

$$P(A) \approx \frac{n_A}{n} \tag{8.2}$$

另外,也可以用**大数定律**(即著名的 **Bernoulli 定理**)表述:如果 n_A 是一列试验中 A 发生的次数,$P(A)$是一次试验中 A 发生的概率,那么,随着 n 增大,相对频数(n_A/n)将愈来愈接近 $P(A)$. 换句话说,试验次数愈大,$P(A)$的相对频数估计愈好. 对于上述的掷硬币例子,100 次试验充分证实你的怀疑:这枚硬币不公正——硬币似乎"被设置"为出现正面多于出现反面.

例 8.2　一张**随机数表**(见 3.23 节)由上千个数组成——每个数是从 0 到 9 的十个数字的任意一个,并且十个数字在表中任意位置等可能地出现. 因此,如果进行试验:从该表随机选取一个数,则可以使用古典解释计算这种选取(一个数或一组数)产生的事件的概率. 例如,对于 P(一个小于 5 的数),因为其中有 5 个数字小于 5(0 至 4),因此 $N_A=5$;并且总共有 10 个可能数字(0 至 9),因此 $N=10$. 于是,由公式(8.1),

$$P(\# < 5) = \frac{N_A}{N} = \frac{5}{10} = 0.50, \quad 或 50\%$$

用附录表 A.1 通过以下操作检验这个概率:以任何方式进入表 A.1,从表内相继取出 240 个一位数. 将每取出一个数看作一次试验. 每 10 次试验后累加频数和试验次数计算相对频数(如最初 10 次试验中小于 5 的数的频数,最初 20 次,最初 30 次,等等),并绘制相对频数图.

解 ☞　我们从第 1 行与第 1 列进入表 A.1 并按行向右前进,从每 5 个相继列取出 10 个数. 这样,从第 1 行的 1 至 5 列取出 10 个数:1,0,0,9,7,3,2,5,3,3;再从第 1 行的 6 至 10 列取出 10 个数:7,6,5,2,0,1,3,5,8,6;等等. 为得到 240 个数,首先从第 1 行的 1 至 75 列选取,然后从第 2 行的 1 至 45 列选取,其中,从第 2 行的 41 至 45 列取出的最后 10 个数是:8,8,6,9,5,4,1,9,9,4. 结果如图 8-1 所示,其中纵轴是相对频数(n_A/n),横轴是试验次数(n),水平虚线表示事件的古典频数[$P(\#<5)=N_A/N=0.50$],用折线连接的实心圆点表示小于 5 的数的相对频数,按每 10 次试验后累积频数和试验次数计算得到.

如大数定律所述,从图 8-1 可以看出,随着试验次数的增加,概率的相对频数估计稳定在古典概率附近. 古典概率和相对频数估计之间的这种一致性,充分证明:表 A.1 确实是一张随机数表,即 10 个可能数字等可能地出现在表中任意位置. 其实,无需对此感到惊异,因为如果读到过有关这张表最初来源的介绍,你就会明白,这张表(RAND 公司,百万随机数,第三版,自由出版社,Glencoe,1995)是通过"电子轮盘赌"产生. 该轮盘赌经过仔细测试和校正专门用于产生随机数表.

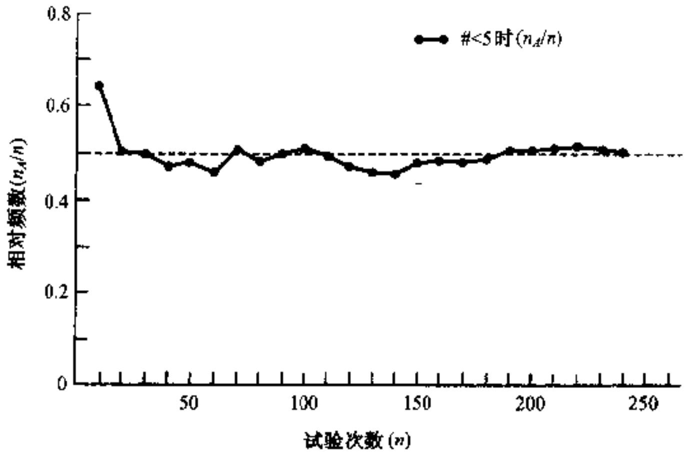


图 8-1

8.3　集合,子集和样本空间

1.17 节指出,集合是一类事物(如物品,符号,数字,单词等)的一个特定聚集. 集合论是研究集合及其特征和相互关系的一个数学分支. 在这里,讨论集合论是因为它是概率的数学理论基础. 不过,我们只论述集合论的有限的几个方面,它们对于直观的理解推断性统计学(见

3.14 节）是必不可少的.

属于或**包含于**一个集合的事物称为该集合的**元素**或**成员**. 表示一个集合既可以**列举**元素，也可以指明元素的**规定性质**. 例如，设一个集合是从 1 到 5 的所有整数. 该集合用列举方法表示为

$$S=\{1,2,3,4,5\}$$

这里，集合的元素列举在花括号内，并且等于符号 S. 用指明规定性质的方法，该集合表示如下

$$S=\{x \mid x\text{是}1\leqslant x\leqslant 5\text{的整数}\}$$

这里，符号 $x|x$（也可写成 $x:x$）表示"满足……的 x."完整的公式可以读成："S 是满足 $1\leqslant x\leqslant 5$ 的整数 x 的集合."以上仅是规定性质的正规表述形式，本书使用比较简单的形式表述集合元素的规定性质. 例如

$$S=\{\text{从 1 到 5 的所有整数}\}$$

如果集合用列举元素的方法表示，则元素的列举次序无关紧要. 例如：历史专业的学生或某学院图书馆收藏的经济类图书，显然没有必要对集合元素进行排序.

子集是集合的一部分. 例如，$S=\{1,2,3,4,5\}$，以下是 S 的一些子集：$A=\{1,2\}$，$B=\{1,2,3\}$，$C=\{1,4,5\}$. 或许很奇怪，全集竟然也可以被视为一个子集，就如根本不包含任何元素的集合也被视为一个子集一样，后者称为**空集**或**零集**，用 $\varnothing$ 表示.

8.1 节讲过，在统计学中，试验是指产生测量值的任何过程. 每个统计试验都有一个**样本空间**——以试验的所有可能结果为元素的集合. 例如，对于 8.1 节的掷骰子试验，样本空间是 $S=\{1,2,3,4,5,6\}$；对于例 8.1 的掷硬币试验，样本空间是 $S=\{$正面，反面$\}$. 注意：用于表示样本空间的符号与表示集合的符号是一样的.

掷骰子试验和掷硬币试验都只有有限多个可能结果，因此，试验的样本空间称为**有限样本空间**. 相反，较高测量水平的试验可以产生无穷多个可能结果，从而构成**无限样本空间**. 例如，试验是以°F（区间间隔水平）测量空调温度，或是以厘米（比例水平）测量孩子的身高，则两个试验是连续型度量，理论上可以产生出无穷多个试验结果. 本章只处理有限样本空间，但后面几章将对二者均有涉及.

8.4　事件

8.1 节指出，来自试验的任何一个明确结果或一组明确结果称为一个事件. 现在，用集合论的语言重新表述这个概念，称样本空间的任何确定子集为一个事件. 一个**简单事件**（或**基本事件**）是这样一个子集，它只包含一个结果（元素）并且该结果不能细分（分解）为更简单更基本的结果. 简单事件的符号是小写字母 e，并带有下标——用以表示它在样本空间中的位置. 于是，对于掷骰子试验，样本空间 $S=\{1,2,3,4,5,6\}$ 有 6 个简单事件：$e_1=\{1\}$，$e_2=\{2\}$，$e_3=\{3\}$，$e_4=\{4\}$，$e_5=\{5\}$，$e_6=\{6\}$.

复合事件也定义为样本空间的一个子集，该子集包含多于一个简单事件. 复合事件常用大写字母表示. 例如，在掷骰子试验中，复合事件 A 可以是：掷出一个偶数点. 这样，A 包含 3 个简单事件：$e_2=\{2\}$，$e_4=\{4\}$，$e_6=\{6\}$，并且 A 可以表示为以下任一种形式：$A=\{e_2,e_4,e_6\}$，$A=\{2,4,6\}$或 $A=\{$掷出一个偶数点$\}$. 一个复合事件发生，如果它所包含的任一个简单事件发生.

由集合论可知，如果一个集合包含 n 个元素，则该集合有 2^n 个子集. 于是，对于掷骰子试验，样本空间 $S=\{1,2,3,4,5,6\}$，$n=6$，则这个集合有（$2^n=2^6=64$）个子集. 每个子集都是来自试验的一个事件，而试验依赖结果的分类方式.

例 8.3　3 个一模一样的球，编号为 1，2 和 3，放在一个广口瓶中. 试验是不进行观察取出一个球. 试问：这个试验能产生多少个事件（结果或成组结果）？

解 ☞　试验的样本空间 $S=\{1,2,3\}$. 由于 $n=3$，故集合有（$2^n=2^3=8$）个子集. 结果之一是取出 #2 球（简单事件），也可以是取出 #2 球或 #3 球（复合事件，表示为$\{2,3\}$）. 所有的 8 个可能事件是：3 个简单事件[$\{1\}$，$\{2\}$，$\{3\}$]；3 个复合事件[$\{1,2\}$，$\{1,3\}$，$\{2,3\}$]，即取出两个指定球的任一

个;复合事件{1,2,3},集合本身作为一个子集,即取出3个球中的任一个;最后一个事件,不取出任何球(∅,空集).(尽管不取出任何球不会是这个试验的实际结果,但是它毕竟是一个子集,并有0概率.)

8.5 Venn图

Venn图用以表示样本空间以及空间中的事件. 构造Venn图有三种典型方法,如图8-2所示,给出的3个图都表示掷骰子试验的样本空间 $S=\{1,2,3,4,5,6\}$. 样本空间通常用封闭图形表示,诸如椭圆[图8-2(a)]或矩形[图8-2(b)和8-2(c)]. 样本空间中的简单事件可以用实心点表示,称为**样本点**[图8-2(a)和8-2(b)],也可以假想它们存在但并不显示[图8-2(c)]. 复合事件用样本空间内的封闭图形表示. 复合事件 $A=\{$掷出一个偶数点$\}$既是图8-2(a)和8-2(b)中阴影椭圆,也是图8-2(c)中的阴影圆形. 如果显示样本点,无论是否标明符号,此时Venn图也称为**Euler图**. 如果不显示样本点,样本空间中封闭图形的面积大小无关紧要;面积与图形所包含的样本点个数不存在比例关系.

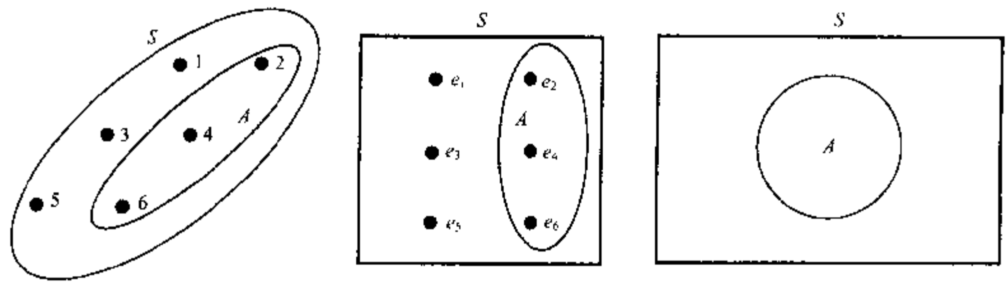


图8-2

样本空间S中,任一事件A的补是S的如下子集:S中不属于A的所有元素. 这个子集用符号 A'(或 $\overline{A}$)表示. 在图8-2(a),(b)和(c)中,虽然 A' 可以视为S中A以外的部分,但是A和 A' 的最常用的Venn图如图8-3所示. 这里,包含A的圆形没有阴影,而 A' 是阴影区域,即圆形之外矩形之内的部分.

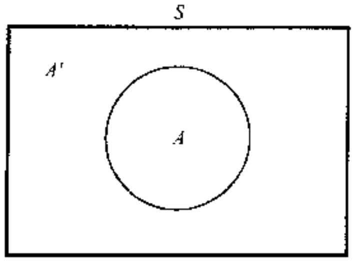


图8-3

两个事件称为**互斥事件**(或**不相交事件**),如果其中一个发生另一个必然不能发生. 用集合论的语言,两个事件互斥,如果它们没有任何共同元素;即不存在这样的元素,它既属于一个事件又属于另一个事件. 根据定义,样本空间中的任意两个简单事件都是互斥的. 例如,掷一枚硬币同时观测到正面和反面是不可能的,或者掷一粒骰子同时观测到1点和6点也是不可能的. 两个复合事件是互斥的,如果它们没有共同的简单事件.

例8.4 用Venn图描述掷骰子试验的样本空间,并图示两个互斥事件:$A=\{e_1,e_3\}$ 和 $B=\{e_4,e_6\}$.

解 互斥事件A和B的Venn图作在图8-4(a),并且显示出了样本点. 描述互斥事件的Venn图更典型的是做成图8-4(b)的形式. 在这两个图中,里面的图形相互分离并且没有阴影.

样本空间S中,两个事件A和B的**并**是S的如下子集:它的元素或者属于A或者属于B,或者既属于A又属于B. 换句话说,试验可以导致A或B或A和B发生. 两个事件的并有以下几种表示方法:$A\cup B$,$A+B$,A或B. 对于掷骰子试验,事件 $A=\{3,4\}$ 和 $B=\{2,4,6\}$ 的并的Venn图如图8-5所示.(a)显示出样本点并且两个椭圆均有阴影. 更典型的形式,用两个都有阴影的事件圆重叠在一起表示A和B的并,见图(b).

样本空间S中,两个事件A和B的**交**是S的如下子集:它的元素既属于A又属于B. 换句话说,试验可以导致A和B同时发生. 两个事件的交有以下几种表示方法:$A\cap B$;A,B;

AB；A 和 B. 对于掷骰子试验，事件 $A=\{3,4\}$ 和 $B=\{2,4,6\}$ 的交的 Venn 图如图 8－6(a) 所示，其中还显示出样本点. 事件 A 和 B 的交的更典型的 Venn 图形式不显示样本点，如图 8－6(b) 所示. 注意：在两个图中，只有相交（重叠）的部分有阴影.

例 8.5　对于掷骰子试验，$S=\{1,2,3,4,5,6\}$，如果 $A=\{1,2\}$ 和 $B=\{2,3,4\}$，求：(a) $A\cup B$，(b) $A\cap B$，(c) A 或 B，(d) A 和 B？

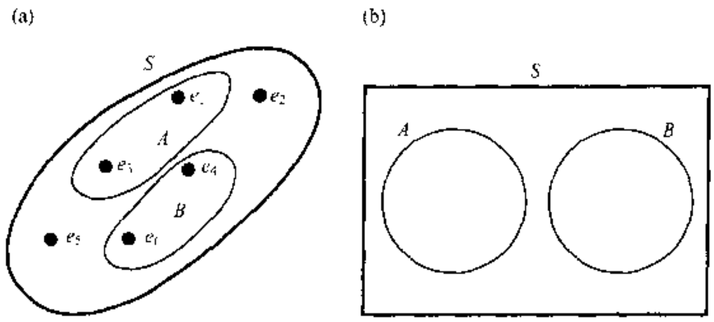


图 8－4

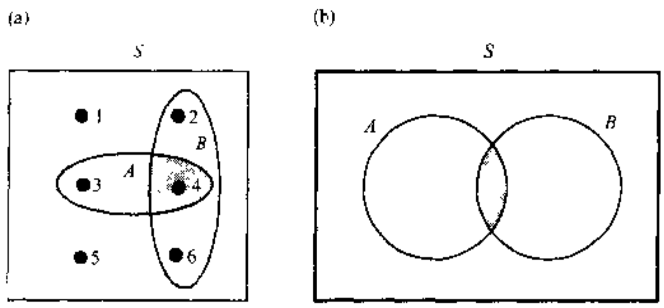


图 8－5

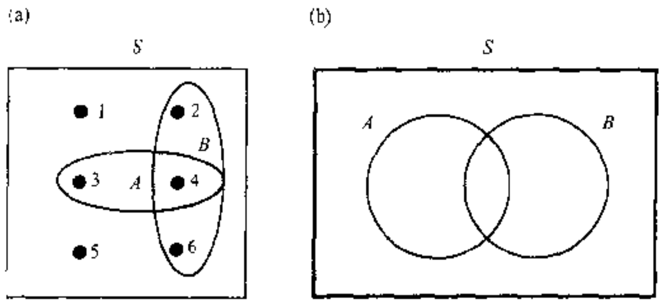


图 8－6

解 ☞　(a) $A\cup B$ 是 A 和 B 的并，因此是 S 的一个子集，它由 S 中属于 A，属于 B，或者既属于 A 又属于 B 的所有元素构成. 于是

$$A\cup B=\{1,2,3,4\}$$

(b) $A\cap B$ 是 A 和 B 的交，因此是 S 的一个子集，它由 S 中既属于 A 又属于 B 的所有元素构成. 于是

$$A\cap B=\{2\}$$

(c) A 或 B 同 $A\cup B$ 一样.

(d) A 和 B 同 $A\cap B$ 一样.

8.6　概率的集合论解释

集合论是概率的数学理论基础. 8.3 到 8.5 节已经讲过一些它的概念和运算. 同其它的数

学理论一样,概率论也有一些给定的并且其正确性毋庸置疑的假设,即所谓的**公理**(或**公设**).理论的所有其它部分都是通过逻辑演绎从公理推导得到.概率论有三个这样的公理,我们将在稍后介绍,概率的所有性质和原理均能由此导出.

在数学理论中,概率与现实世界中的特定现象(如赌博)没有关系,它是完全在理论背景下定义和提出的一个抽象概念,就像点和线是理论几何的抽象概念一样.在这个理论框架中,**概率函数**定义为一个数学函数——使样本空间中的事件对应一个实数(称为概率),且满足三个公理.概率函数以样本空间中的事件作为**定义域**,以指定给这些事件的概率作为**值域**.(见1.17节关于函数的讨论.)到本章为止,已经讨论了两类这样的函数:来自古典解释的函数,$P(A)=N_A/N$,即公式(8.1),称为**古典概率函数**;和来自相对频数解释的函数,$P(A)\approx n_A/n$,公式(8.2),称为**相对频数函数**.

以下是**概率的集合论解释**:

设样本空间S包含有简单事件和复合事件,则这些事件的概率定义为一个函数,使样本空间中的每一事件对应一个确定的实数(称为概率),并且满足以下公理:

公理 I 对S中事件A

$$P(A)\geqslant 0$$

公理 II 对样本空间S

$$P(S)=1$$

公理 III 如果S中事件A和B互斥,则

$$P(A\cup B)=P(A)+P(B)$$

公理 I 表明,事件A的概率总是非负实数;即$P(A)$总大于或等于0.公理 II 表明,试验的每次试验必然有S中的一个事件发生;即S中有一个事件发生的概率是100%.由于公理 II,S称为**必然事件**.公理 III(也称为**特殊加法原理**)表明,两个互斥事件A和B的并的概率等于两个事件的概率和.

尽管上述三个公理是数学抽象的一部分,但是它们作为概率论公理的真正原因在于:从这三个公理出发,可以推导出古典概率和相对频数概率的所有已知性质.本书无意进行这种正规的数学推导,只是简单的给出能由此导出的性质和原理.以下是这样的七个性质:

性质 1 对于S中不可能事件$\varnothing$

$$P(\varnothing)=0$$

性质 2 对于S中事件A

$$0\leqslant P(A)\leqslant 1$$

性质 3 对于事件A和它的补A'

$$P(A)+P(A')=1$$

性质 4 如果S中事件$A_1,A_2,\cdots,A_k$互斥,则

$$P(A_1\cup A_2\cup\cdots\cup A_k)=P(A_1)+P(A_2)+\cdots+P(A_k)$$

性质 5 如果S包含n个简单事件e_i,并且每一个有概率$P(e_i)$,则

$$\sum_{i=1}^{n}P(e_i)=1$$

性质 6 如果S中事件A包含k个简单事件e_i,则

$$P(A)=\sum_{i=1}^{k}P(e_i)$$

性质 7 如果S包含N个等可能的简单事件e_i,则

$$P(e_i)=\frac{1}{N},P(A)=N_A\left(\frac{1}{N}\right)$$

$\varnothing$是S中的空子集或零子集,也称为**空事件**或**零事件**,从而性质1说明空事件的概率为0%.由于这个原因,空事件也称为**不可能事件**,即它永远不会发生.

由公理 I 知,$P(A)$总大于等于 0;由公理 II 知,必然事件 S 的概率是 1. 性质 2 将这两个公理合在一起:所有 $P(A)$值都是从 0 到 1 范围上的非负实数.

A 和 A' 是互斥事件. 因此,由公理 III,

$$P(A \cup A') = P(A) + P(A')$$

我们也知,A'包含集合中不属于 A 的所有元素(见 8.5 节),于是

$$A \cup A' = S$$

因此

$$P(A \cup A') = P(S) = P(A) + P(A') = 1$$

即,性质 3 说明一个事件发生或不发生的概率是 100%. 性质 3 也可以写成

$$P(A) = 1 - P(A')$$

或

$$P(A') = 1 - P(A)$$

性质 4 说明公理 III 可以推广到 S 中的任意多个互斥事件.

S 包含 n 个互斥事件 e_i. 因此

$$S = e_1 \cup e_2 \cup \cdots \cup e_n$$

且

$$P(e_1 \cup e_2 \cup \cdots \cup e_n) = P(S) = 1$$

由性质 4 知

$$P(e_1 \cup e_2 \cup \cdots \cup e_n) = P(e_1) + P(e_2) + \cdots + P(e_n) = \sum_{i=1}^{n} P(e_i)$$

于是

$$\sum_{i=1}^{n} P(e_i) = 1$$

因此,性质 5 说明 S 中所有简单事件的概率和为 100%.

性质 6 说明事件 A 的概率是 A 包含的简单事件的概率和.

最后,性质 7 说明对于特例,即 S 中的所有简单事件均等可能(古典解释的条件),且如果 S 中有 N 个简单事件 e_i,则每一个简单事件的概率为 $1/N$,并且任意复合事件 A 的概率是 A 中简单事件的个数 N_A 乘以 $1/N$。

例 8.6　本节指出,任意数学函数都可以是一个概率函数,如果它既能对 S 中的事件赋予确定的概率实值,又满足三个公理. 用掷骰子试验的例子说明函数 $P(A)=N_A/N$,即公式(8.1),称为古典概率函数,实质上是一个概率函数.

解 ☞　8.1 节指出,对于古典解释适用的任何试验,都可应用 $P(A)=N_A/N$. 在这些条件下,

$$P(A) = \frac{N_A}{N} = \frac{\text{有利于 } A \text{ 的结果个数}}{\text{所有可能结果的总数}}$$

用集合论解释,这可表述为

$$P(A) = \frac{N_A}{N} = \frac{\text{属于 } A \text{ 的简单事件的个数}}{\text{属于 } S \text{ 的简单事件的总数}}$$

显然,这个函数使 S 中的事件对应概率实数,从而满足概率函数定义的第一部分. 为说明它也满足三个公理,我们必须说明:如果用该函数替换公理中的 $P(A)$,公理仍然成立.

公理 I,$P(A)\geqslant 0$,实际上是说明 S 中的每一事件必然有大于等于 0 的概率. 对于掷骰子试验,$S=\{1,2,3,4,5,6\}$,这显然成立. 若 $N_A=0$,这对 S 中的空集成立,例如 $A=\{$掷出一个大于 6 的点数$\}$,则 $P(A)=0$. 对 S 的所有其它子集,$N_A>0$,从而 $P(A)>0$.

公理 II,$P(S)=1$,实际上是说明在每一次试验中,S 中的某事件必然发生. 对于掷骰子试验,显然,每一次掷骰子,骰子着地后必然有一面朝上,从而 S 中的某一简单事件必然发生.

公理 III 表述的是:对于互斥事件 A 和 B 有 $P(A\cup B)=P(A)+P(B)$. 即有以下关系成立:

$$P(A \cup B) = \frac{N_{A\cup B}}{N} = \frac{N_A + N_B}{N} = \frac{N_A}{N} + \frac{N_B}{N} = P(A) + P(B)$$

我们举例说明对掷骰子试验这也成立,用两个互斥事件 $A=\{1,2\}$ 和 $B=\{3,4,5\}$,这样,

$$N_A=2,\quad N_B=3,\quad N_{A\cup B}=5,\quad N=6$$

且

$$\begin{aligned}P(A\cup B)&=\frac{N_{A\cup B}}{N}=\frac{5}{6}\\&=\frac{N_A+N_B}{N}=\frac{2+3}{6}\\&=\frac{N_A}{N}+\frac{N_B}{N}=\frac{2}{6}+\frac{3}{6}\\&=P(A)+P(B)\end{aligned}$$

8.7 概率的主观解释

用古典或相对频数函数确定的概率都称为**客观概率**.这些概率的确定完全来自客观信息:关于事件可能性的明显的实际信息,即没有受到任何个人的感觉或偏见的影响.例如,古典概率是在事先已知试验的所有可能结果,并且它们是等可能出现的条件下确定的,而相对频数概率是在已知事件在一列试验中出现次数的比例的条件下确定的.

然而,也存在许多这种例子,没有必要确定出概率,而且也不可能完全根据客观信息确定概率.这些实例需要进行"主观判断"或"先验猜测",即利用个人对信息的独特判断来确定事件的出现概率.在这些实例中,常用一个数值作为对某一事件的可能性的**信任度**(或**确信度**).这种信任度的度量是**主观的**,从而概率的这种形式称为**概率的主观**(或个性或人为)**解释**.

例如,如果一个试验从未实际进行过,则用主观解释是合适的.假设你是一个商人,将把一种新产品推向市场.为确定产品的成功概率,你首先对可获得的信息(例如,同类产品的成功或失败,你对推广新产品的以往的经验,该市场的一般经济环境)进行了评估,并且凭借感觉和直觉,最后用某种方式将所有的这一切综合成主观概率值:有 0.80 的成功概率.由于这种解释,你在心理上已确信该产品成功的可能 4 倍于失败的可能.

再举一个关于试验的主观概率的实例,一名高中毕业生判断她有 50%的机会被自己理想的院校接受.根据对可利用信息的评估和信念,她认为被接受或被拒绝是等可能的.

通常,主观概率的概率函数无法具体和正规的写出.但是,无论在计算中引入怎样的技巧、经验和心理因素,得到的概率值必须满足 8.6 节的公理和性质以及将要在第 9 章讨论的其它原理和性质.

8.8 机会比率的概念

机会比率的概念是表示概率(无论主观的还是客观的)的又一种方法.如果 $P(A)$是事件 A 发生的概率,而 $P(A')$是事件 A 不发生的概率(它的补的概率),则**有利于事件 A 发生的机会比率**定义为 $P(A)$对 $P(A')$的比.习惯上将这个比表示为两个没有公因子的正整数 c 和 d 的比.于是,有利于 A 的机会比率是

$$\frac{P(A)}{P(A')}=\frac{c}{d}\tag{8.3}$$

通常称为有利于 A 的机会比率是 c 比 d.不利于 A 的机会比率是

$$\frac{P(A')}{P(A)}=\frac{d}{c}\tag{8.4}$$

通常称为不利于 A 的机会比率是 d 比 c.

如果 $P(A)>P(A')$,机会比率习惯的称为有利于 A 的机会比率;如果 $P(A)<P(A')$,称为不利于 A 的机会比率.

例 8.7 电视台的天气预报员称,明天下雨的概率为 70%.根据预报员的说法,明天下雨的机会比率是多少?

解 对于这个问题,事件 A 是明天下雨,并且预报员声称 $P(A)=0.70$.由 8.6 节性质 3 知

$P(A')=1-P(A)$. 因此

$$P(A')=1-0.70=0.30$$

于是，有利于明天下雨的机会比率是

$$\frac{P(A)}{P(A')}=\frac{0.70}{0.30}=\frac{7}{3},\quad \text{或 7 比 3}$$

同商业赌博和赛马给出的定义一样，**赌博机会比率**是就某一事件打赌的赌徒能够获胜或面临失败的一个度量. 例如，赌博机会比率 5 比 1 表示，赌博者将在该赌博中赢得 \$5 或失去 \$1. 博彩公司设置赌博机会比率是为了盈利；因此，针对某一事件的机会比率不会始终保持不变. 如果在某一赌博中这两类机会比率相同，则该赌博称为**公平赌博**.

例 8.8　你已经知道，两个网球选手 R 和 S 已进行过 36 场比赛并且 R 在其中的 24 场获胜. 你以 \$15 对 \$10 和朋友打赌：R 将赢得下一场比赛. 这是公平赌博吗？

解 ☞　$A=\{R\text{ 获胜}\}$，$A'=\{S\text{ 获胜}\}$. 由相对频数解释

$$P(A)\approx\frac{24}{36}=\frac{2}{3}\quad P(A')\approx\frac{12}{36}=\frac{1}{3}$$

因此，有利于 A 的机会比率是

$$\frac{P(A)}{P(A')}\approx\frac{2/3}{1/3}=\frac{2}{1},\quad \text{或近似 2 比 1}$$

于是，如果这个赌博是公平赌博，你和朋友之间应该以 \$20 对 \$10 打赌. 换句话说，这不是公平赌博.

8.9　由机会比率确定概率

在 8.8 节我们是由概率确定机会比率. 在本节，我们由机会比率来确定概率，即通过以下公式解得 $P(A)$ 和 $P(A')$.

$$\frac{P(A)}{P(A')}=\frac{c}{d}$$

由 8.6 节性质 3 知

$$P(A')=1-P(A)$$

将这个关系式代入机会比率公式

$$\frac{P(A)}{1-P(A)}=\frac{c}{d}$$

从而

$$d\times P(A)=c\times[1-P(A)]=c-c\times P(A)$$

因此

$$[c\times P(A)]+[d\times P(A)]=c$$
$$P(A)(c+d)=c$$

从而

$$P(A)=\frac{c}{c+d}\tag{8.5}$$

为解得 $P(A')$，再次利用 8.6 节性质 3

$$P(A')=1-P(A)$$

因此

$$\begin{aligned}P(A')&=1-\frac{c}{c+d}\\&=\frac{c+d}{c+d}-\frac{c}{c+d}\\&=\frac{c+d-c}{c+d}\end{aligned}$$

从而

$$P(A') = \frac{d}{c+d} \tag{8.6}$$

例 8.9　如果有利于 A 发生的机会比率是 4 比 3,试问:(a)$P(A)$,(b)$P(A')$是多少?

解 ☞　由于 $c=4$ 和 $d=3$:

(a)$P(A)=\frac{c}{c+d}=\frac{4}{4+3}=0.57$,或 57%

(b)$P(A')=\frac{d}{c+d}=\frac{3}{4+3}=0.43$,或 43%

习题解答

概率的古典解释

8.1　一副标准的 52 张扑克牌有 4 种花色(梅花,方块,红心和黑桃),每一花色有 13 张牌(A,2,3,4,5,6,7,8,9,10,J,Q 和 K).方块和红心称为红色牌(标有红色符号),梅花和黑桃称为黑色牌(标有黑色符号).试验是从一副完全洗乱的扑克牌中选取一张.计算以下概率:(a)P(红心 10),(b)P(红心),(c)P(10),(d)P(红色牌).

解 ☞　(a)因为取出 52 张牌的任一张都是等可能的,故 $N=52$.并且 52 张牌中仅有一张有利于该事件,故 $N_A=1$.因此,由公式(8.1),

$$P(\text{红心 }10)=\frac{N_A}{N}=\frac{1}{52}=0.019,\quad \text{或 }1.9\%$$

(b)再次有 $N=52$,但是现在有 13 种可能的有利结果(红心),从而 $N_A=13$.因此

$$P(\text{红心})=\frac{N_A}{N}=\frac{13}{52}=0.25,\quad \text{或 }25\%$$

(c)再次有 $N=52$,但是现在有 4 种可能的有利结果(牌面标有 10).从而 $N_A=4$.因此

$$P(10)=\frac{N_A}{N}=\frac{4}{52}=0.077,\quad \text{或 }7.7\%$$

(d)再次有 $N=52$,但是现在有 26 种可能的有利结果(方块和红心),从而 $N_A=26$.因此

$$P(\text{红色牌})=\frac{N_A}{N}=\frac{26}{52}=0.50,\quad \text{或 }50\%$$

8.2　表 8.1 给出某学院历史专业的 85 名学生的年龄和性别(如有 15 名学生为男生且年龄为 20 岁或更小).试验是使用一张随机数表(见 3.23 节)从该专业选择一名学生,记录他或她的性别和年龄.计算以下概率:(a)P(年龄为 20 或更小的男生),(b)P(男生),(c)P(年龄为 20 或更小的学生),(d)P(男生或女生).

表 8.1

	年龄		总计
性别	20 或更小	超过 20	
男生	15	30	45
女生	20	20	40
总计	35	50	85

解 ☞　(a)因为这个试验的等可能结果是 85 名学生,故 $N=85$,结果等可能是因为选取是随机的.因为有 15 名年龄为 20 或更小的男生,故 $N_A=15$.因此,由公式(8.1),

$$P(\text{年龄为 20 或更小的男生})=\frac{N_A}{N}=\frac{15}{85}=0.18,\quad \text{或 }18\%$$

(b)再次有 $N=85$,但是现在总共有 45 名男生,从而 $N_A=45$.因此

$$P(\text{男生})=\frac{N_A}{N}=\frac{45}{85}=0.53,\quad \text{或 }53\%$$

(c)再次有 $N=85$,但是现在总共有 35 名学生年龄为 20 或更小,从而 $N_A=35$.因此

$$P(\text{年龄为 20 或更小的学生})=\frac{N_A}{N}=\frac{35}{85}=0.41,\quad \text{或 }41\%$$

(d)再次有 $N=85$,但是现在所有学生要么为男生要么为女生,从而 $N_A=85$.因此

$$P(\text{男生或女生})=\frac{N_A}{N}=\frac{85}{85}=1.0,\quad \text{或 }100\%$$

8.3 50粒不同颜色的石子放入一只瓶子并且完全混合在一起.石子中有25粒蓝色,20粒绿色和5粒红色.如果闭上眼睛从瓶中取出一粒石子,计算以下概率:(a)P(红色石子),(b)P(蓝色或红色石子).

解 ☞ (a)$N=50$,$N_A=5$.由公式(8.1),

$$P(\text{红色石子})=\frac{N_A}{N}=\frac{5}{50}=0.10,\quad \text{或}10\%$$

(b)$N=50$,$N_A=25+5=30$.由公式(8.1),

$$P(\text{蓝色或红色石子})=\frac{N_A}{N}=\frac{30}{50}=0.60,\quad \text{或}60\%$$

8.4 三个家庭一起度假进餐.Brown夫妇有3个女儿和1个儿子.Cruz夫妇有1个女儿和2个儿子.Hansen夫妇有3个儿子.为了确定谁将切火鸡,每个人将自己的名字写在一张纸片上并放入帽子中.纸片在帽子中完全混合,不进行观察,Hansen先生随意从帽子中取出一个名字.试计算取出以下名字的概率:(a) Hansen先生,(b) Cruz家庭的一个成员,(c)一个男性,(d) 非Hansen家庭的一个成员.

解 ☞ (a)$N=6+5+5=16$, $N_A=1$.由公式(8.1),

$$P(\text{Hansen 先生})=\frac{N_A}{N}=\frac{1}{16}=0.06,\quad \text{或}6\%$$

(b)$N=6+5+5=16$,$N_A=5$

$$P(\text{Cruz 家庭的一个成员})=\frac{N_A}{N}=\frac{5}{16}=0.31,\quad \text{或}31\%$$

(c)$N=6+5+5=16$,$N_A=2+3+4=9$

$$P(\text{一个男性})=\frac{N_A}{N}=\frac{9}{16}=0.56,\quad \text{或}56\%$$

(d)$N=6+5+5=16$,$N_A=6+5=11$

$$P(\text{非 Hansen 家庭的一个成员})=\frac{N_A}{N}=\frac{11}{16}=0.69,\quad \text{或}69\%$$

概率的相对频数解释

8.5 在8.1节,古典解释给出掷骰子试验的两个可能事件的概率:P(掷出3点)$=0.17$,P(掷出偶数点)$=0.50$.找一粒骰子,连续掷240次来检验这两个概率.要求使各次试验尽可能相似.根据投掷结果,通过在每12次试验后累加频数和试验次数,计算每一事件的相对频数(例如,在最初的12次,24次,36次试验中掷出3点的相对频数,等等.),然后绘制这些相对频数的图形.

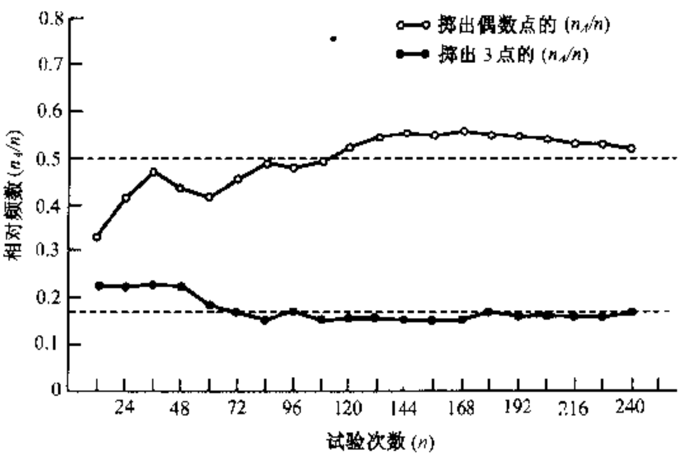


图 8-7

解 ☞ 掷一粒标准的塑料骰子240次,用"相同的动作":将骰子掷到离墙大约4英尺的地方,然后让

骰子自由滚动直至停下.掷出偶数点和掷出3点的概率如图8-7所示,这里,纵轴是相对频数(n_A/n);横轴是试验次数(n);两条水平虚线表示两个事件的古典概率(N_A/N);连接实心圆的折线表示每12次试验后累加得到3点的相对频数;连接空心圆的折线表示每12次试验后累加得到偶数点的相对频数.

可以看出,同大数定律(见8.2节)预言的一样,随着试验次数的增加,累积相对频数变得愈来愈稳定并且愈来愈接近这些事件的古典概率.如果将试验继续下去,相对频数将精确的落在古典概率直线上,或者由于试验方法或骰子的原因,与古典概率直线有稍微的偏离.

8.6 在某大城市一家医院的产妇病房,去年出生1060个男婴和1000个女婴.假设这些数据表示了全部出生情况,在该医院下一个出生的婴儿是男婴的概率是多少?是女婴的概率是多少?

解☞ 对男婴,$n_A=1060$ 且 $n=(1060+1000)=2060$.于是,由公式(8.2),

$$P(\text{男婴})\approx\frac{n_A}{n}=\frac{1060}{2060}=0.51,\quad \text{或 } 51\%$$

对女婴,$n_A=1000$ 且 $n=2060$.于是,

$$P(\text{女婴})\approx\frac{n_A}{n}=\frac{1000}{2060}=0.49,\quad \text{或 } 49\%$$

8.7 在赛季初期(五月下旬),一个棒球选手击球26次只击中5次.在本赛季末(八月上旬),他击球352次击中117次.在(a)五月下旬,(b)八月上旬,他下次击中的概率是多少?

解☞ (a)在五月下旬,他击中的相对频数(平均击中率)是5/26=0.19.因此,他下次击中的概率的相对频数估计是

$$P(\text{击中})\approx\frac{n_A}{n}=0.19$$

(b)在八月上旬,相对频数估计是

$$P(\text{击中})\approx\frac{n_A}{n}=\frac{117}{352}=0.33$$

由于八月的试验次数大约是五月的试验次数的14倍,因此,由大数定律,认为八月份的估计是他击中概率的一个更好的估计.

8.8 一个汽车前灯制造商想知道公司生产的下一个汽车前灯是不合格品的概率.他检验了100个样品并发现有2个不合格品.试问:下一个是不合格品的概率是多少?在1000个前灯的批中,他期望发现多少个不合格品?

解☞ 对100个样品,$n_A=2$ 且 $n=100$.因此,下一个前灯是不合格品的概率是

$$P(\text{不合格品})\approx\frac{n_A}{n}=\frac{2}{100}=0.02,\quad \text{或 } 2\%$$

由此可以估计,在1000个前灯的一批中有2%或20个不合格品.当然,这并不是一个精确的估计,它仅仅是基于100次试验后获得的一个近似估计.

集合,子集,样本空间和事件

8.9 在机会对策中,从一副完全洗乱的52张标准扑克牌中选取出一张.试问:样本空间是什么,试验将产生多少个事件?

解☞ 有52个可能结果,不使用列举方法,而是使用规定性质的方法表述为:$S=\{$一副标准的52张扑克牌$\}$.于是,由8.4节知,试验将产生 $2^n=2^{52}$ 个可能事件.

8.10 试验是从表8.1列出的85名学生中随机选出一名.试问:样本空间是什么,试验将产生多少个事件?

解☞ 使用规定性质,$S=\{$历史专业的85名学生$\}$,试验将产生 2^{85} 个可能事件.

8.11 试验是掷2次硬币并且每次观测朝上的一面是正面还是反面.试问:样本空间是什么,试验将产生出多少个事件?

解☞ 对于掷硬币试验的2次连续试验,每一简单事件都由两部分组成.用 H 表示正面,T 表示反

面,样本空间 $S=\{HH,HT,TH,TT\}$. 于是,试验将产生($2^4=16$)个可能事件. 这些事件是:

$$\{HH\},\{HT\},\{TH\},\{TT\},\{HH,HT\},\{HH,TH\},\{HH,TT\},$$
$$\{HT,TH\},\{HT,TT\},\{TH,TT\},\{HH,HT,TH\},\{HH,HT,TT\},$$
$$\{HH,TH,TT\},\{HT,TH,TT\},\{HH,HT,TH,TT\} \text{ 和 } \varnothing$$

(回顾 8.4 节知,一个复合事件,例如$\{HH,HT\}$,只能包含其中的一个简单事件,例如 HH.)

8.12 对于习题 8.6 的产妇病房,试验是确定下一个在该病房出生的婴儿的性别. 对于该试验,样本空间是什么,且将产生出多少个事件?

解 ☞ $S=\{$男婴,女婴$\}$;($2^2=4$)个可能事件是:{男婴},{女婴},{男婴,女婴}和$\varnothing$.

8.13 对习题 8.7 的棒球选手,试验是确定他下次击球是否能击中. 对于该试验,样本空间是什么,且将产生出多少个事件?

解 ☞ $S=\{$击中,未击中$\}$;($2^2=4$)个可能事件是:{击中},{未击中},{击中,未击中}和$\varnothing$.

8.14 对于习题 8.8 的汽车前灯制造商,试验是确定下一个前灯是否是不合格品. 对于该试验,样本空间是什么,且将产生出多少个事件?

解 ☞ $S=\{$不合格品,非不合格品$\}$;($2^2=4$)个可能事件是:{不合格品},{合格品},{不合格品,合格品}和$\varnothing$.

Venn 图

8.15 从一副完全洗乱的标准扑克牌中取出一张. 对这个试验,指出以下每组事件是否互斥或者互补:(a)$A=\{$一张红色牌$\}$,$B=\{$一张黑色牌$\}$;(b)$A=\{$牌面为 2,3,4,6,或 10 的一张牌$\}$,$B=\{$一张方块$\}$;(c)$A=\{$牌面为 2,3,4,5,6,7,8,9,或 10 的一张牌$\}$,$B=\{$牌面为 A,J,Q,或 K 的一张牌$\}$;(d)$A=\{$牌面为 2,3,4,5,6,或 7 的一张方块$\}$,$B=\{$牌面为 8,9,10,J,Q,或 K 的一张方块$\}$.

解 ☞ (a)事件 A 和 B 互斥,因为任何一张牌都不可能既是红色又是黑色. A 和 B 也互补($B=A'$),因为 S 中所有不属于 A 的简单事件都属于 B.

(b)事件 A 和 B 不互斥,因为它们有共同的简单事件:牌面为 2,3,4,6,或 10 的方块. 因为不互斥,事件 A 和 B 也自然不互补. 由前知,A'包含样本空间中所有**不属于** A 的简单事件.

(c)事件 A 和 B 互斥:它们没有共同的简单事件. 事件 A 和 B 也互补($B=A'$),因为 S 中所有不属于 A 的简单事件都属于 B.

(d)事件 A 和 B 互斥;方块牌不可能既是 2,3,4,5,6,或 7 又是 8,9,10,J,Q,或 K. 然而,事件 A 和 B 不互补,因为样本空间中其它三种花色不包含于 A 或 B 中.

8.16 从表 8.1 给出的历史专业的班级中随机选取出一名学生. 对这个试验,指明以下每组事件是否互斥或者互补:(a)$A=\{$年龄为 20 或更小的男生$\}$,$B=\{$年龄超过 20 的男生$\}$;(b)$A=\{$男生$\}$,$B=\{$女生$\}$;(c)$A=\{$年龄为 20 或更小的学生$\}$,$B=\{$年龄超过 20 的学生$\}$;(d)$A=\{$年龄为 20 或更小的男生$\}$,$B=\{$年龄为 20 或更小的女生$\}$.

解 ☞ (a)A 和 B 互斥但不互补.

(b)A 和 B 既互斥又互补.

(c)A 和 B 既互斥又互补.

(d)A 和 B 互斥但不互补.

8.17 对于图 8-8 中 Venn 图给出的样本空间,其中包含事件 A,B 和 C,确定以下事件:(a) A',(b)$A\cup A'$,(c)A 和 A',(d)$A\cap B$,(e)A 或 B,(f)B 和 C,(g)$B\cup C$,(h)$A'\cap B$,(i) $A'\cup B$,(j)A'或 B',(k)$B'\cup C'$,(l)B'和C'.

解 ☞ (a)A'是 A 的补,因此包含 S 中不属于 A 的所有元素. 于是

$$A'=\{e_1,e_4,e_6,e_7,e_8,e_9,e_{10},e_{11},e_{12}\}$$

(b)$A\cup A'$,即 A 和它的补的并,是属于 A 或者属于 A'或者既属于 A 又属于 A'的所有元素构成的 S 的一个子集. 于是

$$A \cup A' = S$$

(c) A 和 A'，即 A 和它的补的交，是包含于 A 和 A' 的所有公共元素构成的 S 的一个子集. 由定义，不存在公共元素，故

$$A \text{ 和 } A' = \varnothing$$

(d)事件 A 和 B 没有任何公共元素，因为它们互斥. 于是

$$A \cap B = \varnothing$$

(e) A 或 B 是 A 和 B 的并. 于是

$$A \text{ 或 } B = \{e_2, e_3, e_5, e_6, e_7, e_8\}$$

(f)事件 B 和 C 有公共元素 e_6. 于是

$$B \text{ 和 } C = \{e_6\}$$

S

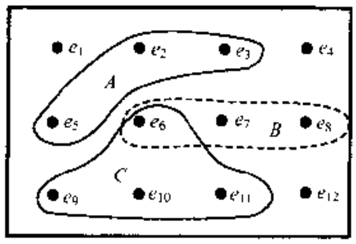


图 8-8

(g) $B \cup C = \{e_6, e_7, e_8, e_9, e_{10}, e_{11}\}$

注意：尽管 e_6 既属于 B 又属于 C，但在 $B \cup C$ 中只能包含一次.

(h) $A' \cap B = \{e_6, e_7, e_8\}$

(i) $A' \cup B = A' = \{e_1, e_4, e_6, e_7, e_8, e_9, e_{10}, e_{11}, e_{12}\}$

(j) A' 或 $B' = S$

(k) $B' \cup C' = \{S$ 中除去 e_6 的所有元素$\}$

(l) B' 和 $C' = \{e_1, e_2, e_3, e_4, e_5, e_{12}\}$

8.18 从一副完全洗乱的标准扑克牌中抽取一张，其中 $S=\{$一副标准的 52 张扑克牌$\}$. 对该试验，考虑以下事件：$A=\{\text{A}\}$，$B=\{$方块$\}$，$C=\{2,3,4,5,$或$6\}$，$D=\{$黑色牌$\}$. 确定以下事件：(a) B'，(b) $A\cup B$，(c) A 和 B，(d) $A\cap B'$，(e) $B\cap D'$，(f) B 或 C，(g) $A\cap C$，(h) C 或 D，(i) A 和 C'，(j) $A'\cap B'$，(k) $B'\cup D'$.

解 ☞ (a) $B'=\{$梅花，红心和黑桃花色的 39 张牌$\}$

(b) $A\cup B=\{$13 张方块和梅花 A、红心 A 和黑桃 A$\}$

(c) A 和 $B=\{$方块 A$\}$

(d) $A\cap B'=\{$梅花 A、红心 A 和黑桃 A$\}$

(e) $B\cap D'=\{$13 张方块$\}$

(f) B 或 $C=\{$13 张方块和其它花色的 2,3,4,5 和 6$\}$

(g) $A\cap C=\varnothing$

(h) C 或 $D=\{$26 张黑色牌(梅花和黑桃)和红色牌(方块和红心)的 2,3,4,5 和 6$\}$

(i) A 和 $C'=\{$4 个 A$\}$

(j) $A'\cap B'=\{$除去 A 以外的所有梅花，红心和黑桃$\}$

(k) $B'\cup D'=S$.

概率的集合论解释

8.19 晚会的票价为 \$4，\$8，\$12，\$15，或 \$20. 前 1000 个观众的购票情况如下：200 个购 \$4 票，500 个购 \$8 票，150 个购 \$12 票，100 个购 \$15 票和 50 个购 \$20 票. 试验是预测下一张被购票的价格. 以该试验为例，说明公式(8.2)，$P(A)\approx n_A/n$，即所谓的相对频数概率函数，实际上是一个概率函数. 以例 8.6 为范例.

解 ☞ 就相对频数解释而言，在一个试验的样本空间中，事件的概率是根据先前进行的一列试验得到的对应事件的相对频数来确定. 于是，与 8.2 节所述相同：

$P(A)$ 近似等于，在一列试验中 A 发生的次数(n_A)与这列试验的总次数(n)的比. 或者，用符号表示为

$$P(A)\approx\frac{n_A}{n}$$

这个函数将 S 中的事件对应于概率实数，满足相对频数函数定义的前半部分. 它也显然满足公理 I，因为 n_A 永远不可能小于 0，从而 $P(A)$ 永远不可能小于 0. 对 S 中的空集(如购一张 \$25 票)，或者当 n_A 为 0(如无人购 \$4 票)时，$P(A)$ 为 0.

公理 II 满足，因为对每一次试验，S 中的某一简单事件必然发生；并且每次购票的价格必然是 \$4，\$8，\$12，\$15，或 \$20.

与例 8.6 一样,可以表明公理Ⅲ也满足.设 $A=\{票价\leqslant \$8\}$和 $B=\{票价\geqslant \$15\}$,则

$$n_A=700,\quad n_B=150,\quad n_{A\cup B}=850,\quad n=1000$$

并且

$$\begin{aligned}P(A\cup B)&\approx\frac{n_{A\cup B}}{n}=\frac{850}{1000}\\&\approx\frac{n_A+n_B}{n}=\frac{700+150}{1000}\\&\approx\frac{n_A}{n}+\frac{n_B}{n}=\frac{700}{1000}+\frac{150}{1000}\\&\approx P(A)+P(B)\end{aligned}$$

8.20 图 8-9 所示的样本空间中包含事件 A,B,C 和 D,每一样本点表示一个等可能(等概率)的简单事件.用 8.6 节的公理和性质计算以下概率:(a) $P(A)$,(b) $P(B)$,(c) $P(C)$,(d) $P(D)$,(e) $P(A\cup C)$,(f) $P(A')$,(g) $P(A\cup B\cup C\cup D)$,(h) $P(A\cap B)$,(i) $P(A\cup C')$.

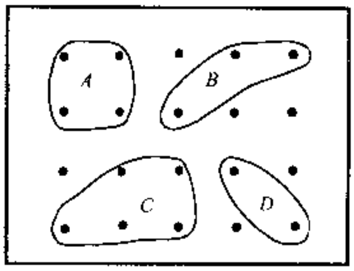


图 8-9

解 ☞　(a)性质 7 说明,如果 S 包含 N 个等可能的简单事件 e_i,则

$$P(e_i)=\frac{1}{N}$$

对这个样本空间,$N=20$,因此

$$P(e_i)=\frac{1}{20}$$

性质 7 进一步表示

$$P(A)=N_A\left(\frac{1}{N}\right)$$

所以,由于 A 中有 4 个简单事件,

$$P(A)=N_A\left(\frac{1}{N}\right)=4\left(\frac{1}{20}\right)=\frac{4}{20}=0.20$$

(b) $P(B)=N_B\left(\frac{1}{N}\right)=3\left(\frac{1}{20}\right)=\frac{3}{20}=0.15$

(c) $P(C)=N_C\left(\frac{1}{N}\right)=5\left(\frac{1}{20}\right)=\frac{5}{20}=0.25$

(d) $P(D)=N_D\left(\frac{1}{N}\right)=2\left(\frac{1}{20}\right)=\frac{2}{20}=0.10$

(e) A 和 C 互斥,因此由公理Ⅲ知

$$P(A\cup C)=P(A)+P(C)=0.20+0.25=0.45$$

(f)由性质 3

$$P(A')=1-P(A)=1-0.20=0.80$$

(g) A,B,C 和 D 均互斥,因此由性质 4 知

$$\begin{aligned}P(A\cup B\cup C\cup D)&=P(A)+P(B)+P(C)+P(D)\\&=0.20+0.15+0.25+0.10=0.70\end{aligned}$$

(h)由于 A 和 B 互斥,$A\cap B=\varnothing$,因此由性质 1 知

$$P(A\cap B)=P(\varnothing)=0$$

(i) A 和 C' 不互斥,因此不能使用公理Ⅲ.由性质 7:

$$A\cup C'=\{S 中不属于 C 的所有简单事件\}$$

从而

$$N_{A\cup C'}=15$$

并且

$$P(A\cup C')=N_{A\cup C'}\left(\frac{1}{N}\right)=15\left(\frac{1}{20}\right)=\frac{15}{20}=0.75$$

8.21 图 8-10 所示的样本空间中包含事件 A、B 和 C，每一样本点上方的数字是样本点表示的简单事件的概率的相对频数近似值 $P(e_i)\approx\frac{n_{e_i}}{n}$. 用 8.6 节的公理和性质计算以下概率：(a)$P(A)$，(b)$P(B)$，(c)$P(C)$，(d)$P(A\cup B)$，(e)$P(B')$，(f)$P(A\cup B\cup C)$，(g)$P(B\cap C)$，(h)$P(A\cup B')$.

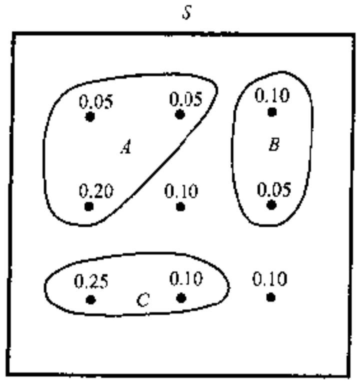


图 8-10

解 ☞　(a)性质 6 表明

$$P(A)=\sum_{i=1}^{k}P(e_i)$$

因此，在这里有

$$P(A)\approx 0.05+0.05+0.20=0.30$$

(b)$P(A)\approx 0.10+0.05=0.15$

(c)$P(A)\approx 0.25+0.10=0.35$

(d)A 和 B 互斥，因此由公理Ⅲ知

$$P(A\cup B)=P(A)+P(B)\approx 0.30+0.15=0.45$$

(e)由性质 3

$$P(B')=1-P(B)$$

因此

$$P(B')\approx 1-0.15=0.85$$

(f)A，B 和 C 均互斥，因此由性质 4 知

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)$$

且

$$P(A\cup B\cup C)\approx 0.30+0.15+0.35=0.80$$

(g)由于 B 和 C 互斥，$B\cap C=\varnothing$，因此由性质 1 知

$$P(B\cap C)=P(\varnothing)=0$$

(h)A 和 B' 不互斥，因此不能使用公理Ⅲ. 由性质 6：

$$A\cup B'=\{S\text{ 中不属于 }B\text{ 的所有简单事件}\}$$

于是

$$P(A\cup B')\approx 0.05+0.05+0.20+0.10+0.25-0.10+0.10=0.85$$

8.22 从一副完全洗乱的标准扑克牌中抽取一张，其中 $S=\{$一副标准的 52 张扑克牌$\}$. 对于这个试验，用 8.6 节的公理和性质计算以下概率：(a)P(梅花 10)，(b)P(K)，(c)P(不是 K)，(d)P(红心或 K)，(e)P(4 或 10 或 K).

解 ☞　(a)由性质 7

$$P(e_i)=\frac{1}{N}=\frac{1}{52}$$

因此，由于仅有一张梅花 10，$N_A=1$ 且

$$P(\text{梅花 }10)=N_A\left(\frac{1}{N}\right)=1\left(\frac{1}{52}\right)=\frac{1}{52}=0.019$$

(b)由性质 7，$N_A=4$ 且

$$P(K)=N_A\left(\frac{1}{N}\right)=4\left(\frac{1}{52}\right)=\frac{4}{52}=0.077$$

(c)由性质 3

$$P(\text{不是 }K)=1-P(K)=1-0.077=0.923$$

(d)$\{$红心或 K$\}=\{$13 张红心以及梅花 K，方块 K 和黑桃 K$\}$. 由性质 7，

$$P(\text{红心或 }K)=N_A\left(\frac{1}{N}\right)=16\left(\frac{1}{52}\right)=\frac{16}{52}=0.308$$

(e)由于 4 或 10 或 K 是互斥的，可以使用性质 4：

$$P(4\text{ 或 }10\text{ 或 }K)=P(4)+P(10)+P(K)$$

由于

$$P(4)=P(10)=P(K)=N_A\left(\frac{1}{N}\right)=4\left(\frac{1}{52}\right)=\frac{4}{52}=0.077$$

则

$$P(4\text{或}10\text{或}K)=0.077+0.077+0.077=0.231$$

8.23 由习题 8.11 的解答,掷两次硬币的试验能产生 16 个可能事件(结果和成组结果).假设确实做了这个试验,并且结果是 HH,则有多少个事件出现并且每一个的概率是多少?

解 ☞ 在试验前,可能的事件有($2^n=2^4=16$)个.实际操作后有 2^{n-1} 个事件出现,因为有这么多个 S 的子集包含试验的结果.这里有($2^{4-1}=2^3=8$)个事件包含 HH:$\{HH\}$,$\{HH,HT\}$,$\{HH,TH\}$,$\{HH,TT\}$,$\{HH,HT,TH\}$,$\{HH,HT,TT\}$,$\{HH,TH,TT\}$和$\{HH,HT,TH,TT\}$.

可以使用 8.6 节的性质 7 计算概率.由于 $S=\{HH,HT,TH,TT\}$,$N=4$,因此

$$P(HH)=N_A\left(\frac{1}{N}\right)=1\left(\frac{1}{4}\right)=\frac{1}{4}=0.25$$

$$P(HH,HT)=P(HH,TH)=P(HH,TT)$$

$$=N_A\left(\frac{1}{N}\right)=2\left(\frac{1}{4}\right)=\frac{2}{4}=\frac{1}{2}=0.50$$

$$P(HH,HT,TH)=P(HH,HT,TT)=P(HH,TH,TT)$$

$$=N_A\left(\frac{1}{N}\right)=3\left(\frac{1}{4}\right)=\frac{3}{4}=0.75$$

$$P(HH,HT,TH,TT)=N_A\left(\frac{1}{N}\right)=4\left(\frac{1}{4}\right)=1$$

概率的主观解释和机会比率的概念

8.24 对以下概率,首先指出是古典,相对频数或主观概率的哪一种;然后,使用 8.8 节的恰当公式将概率变换为机会比率:(a)你估计求职得到工作的概率是 0.10,(b)掷一次骰子不能得到 3 点的概率是 5/6,(c)基于产妇病房去年的男婴和女婴的出生数,该医院下一个出生的婴儿将是男婴的概率近似为 0.51.

解 ☞ (a)主观概率:

$$A=\{\text{得到工作}\},\quad A'=\{\text{不能得到工作}\},\quad P(A)=0.10,\quad P(A')=0.90$$

由公式(8.4),

$$\frac{P(A')}{P(A)}=\frac{0.90}{0.10}=\frac{9}{1}$$

因为 $P(A)<P(A')$,通常说成:你不能得到工作的机会比率是 9 比 1.

(b)古典概率:

$$A=\{3\},\quad A'=\{1,2,4,5,6\},\quad P(A)=1/6,\quad P(A')=5/6,$$

由公式(8.4),

$$\frac{P(A')}{P(A)}=\frac{5/6}{1/6}=\frac{5}{1}$$

不能得到 3 点的机会比率是 5 比 1.

(c)相对频数概率:

$$A=\{\text{男婴}\},\quad A'=\{\text{女婴}\},\quad P(A)\approx 0.51,\quad P(A')\approx 0.49,$$

由公式(8.4),

$$\frac{P(A')}{P(A)}\approx\frac{0.51}{0.49}=\frac{51}{49}$$

从而,有利于下一个婴儿是男婴的机会比率近似为 51 比 49.

8.25 对以下机会比率,首先使用 8.9 节的恰当公式将机会比率变换为概率;然后,指出得到的概率是否为古典,相对频数或主观概率:(a)购买一张 6 位数的彩票,其号码与抽奖号码相同而中彩的机会比率是 5.2 百万比 1;(b)你估计下一个潜在的购买者买下你的房产的机会比率是 3 比 2;(c)从历史数据估计,一个高中生运动员成为职业运动员的机会比率是 500000 比 1.

解 ☞　(a) $A=\{中彩\}$，$A'=\{未中彩\}$，$c=1$，$d=5200000$. 由公式(8.6)，

$$P(A')=\frac{d}{c+d}=\frac{5200000}{1+5200000}=0.99999981,\text{或实际上}100\%$$

古典概率

(b)$A=\{购买\}$，$A'=\{不购买\}$，$c=3$，$d=2$. 由公式(8.5)，

$$P(A)=\frac{c}{c+d}=\frac{3}{3+2}=\frac{3}{5}=0.60$$

主观概率

(c)$A=\{成为职业运动员\}$，$A'=\{不会成为职业运动员\}$，$c\approx1$，$d\approx500000$. 由公式(8.6)，

$$P(A')=\frac{d}{c+d}\approx\frac{500000}{1+500000}=0.9999980,\text{或实际上}100\%$$

相对频数概率

8.26　一个不动产开发商对一处商业住宅评估：在两年之后价值上涨的机会比率是3比2；在两年之后价值保持不变的机会比率是7比3；价值在两年之后下跌的机会比率是3比3. 你认为以上机会比率有错误吗？

解 ☞　将机会比率变换为概率，使用公式(8.5)计算这些概率：

$A_1=\{价值上涨\}$，$A'_1=\{价值不上涨\}$，$c=2$，$d=3$

$$P(A_1)=\frac{c}{c+d}=\frac{2}{2+3}=\frac{2}{5}=0.40$$

$A_2=\{价值保持不变\}$，$A'_2=\{价值不会保持不变\}$，$c=3$，$d=7$

$$P(A_2)=\frac{c}{c+d}=\frac{3}{3+7}=\frac{3}{10}=0.30$$

$A_3=\{价值下跌\}$，$A'_3=\{价值不下跌\}$，$c=2$，$d=3$

$$P(A_3)=\frac{c}{c+d}=\frac{2}{2+3}=\frac{2}{5}=0.40$$

A_1，A_2 和 A_3 互斥. 因此，由8.6节性质4，

$$\begin{aligned}P(A_1\cup A_2\cup A_3)&=P(A_1)+P(A_2)+P(A_3)\\&=0.40+0.30+0.40=1.10\end{aligned}$$

显然，这个结果违背了8.6节性质2，$0\leqslant P(A)\leqslant1$，从而违背了主观概率必须与概率的所有公理和性质保持一致的原则. 因此，开发商应该重新评估这些概率，直到它们一致为止.

8.27　某学生迄今共学习了16门大学课程，并且已通过其中的12门. 该学生用$25和母亲赌$10，认为自己能够通过下学期开设的统计学课程. 这是公平打赌吗？

解 ☞　$A=\{学生获胜\}$，$A'=\{母亲获胜\}$. 由公式(8.2)描述的相对频数解释，有

$$P(A)\approx\frac{n_A}{n}$$

$$P(A)\approx\frac{12}{16}=\frac{3}{4}\quad P(A')\approx\frac{4}{16}=\frac{1}{4}$$

因此，由公式(8.3)，有利于 A 的机会比率是

$$\frac{P(A)}{P(A')}\approx\frac{3/4}{1/4}=\frac{3}{1},\text{或近似3比1}$$

如果这是公平打赌(见例8.8)，应该是学生以$30与母亲的$10作赌，即这不是一个公平打赌.

8.28　两支篮球队每年交手10次. 在过去的8年中，A队获胜48次，B队获胜32次. A队的一个球迷用$30与B队的一个球迷赌$20，声称A队将赢得下次比赛. 这是一个公平打赌吗？

解 ☞　$A=\{A队获胜\}$，$A'=\{B队获胜\}$. 由公式(8.2)描述的相对频数解释，有

$$P(A)\approx\frac{n_A}{n}$$

$$P(A)\approx\frac{48}{80}=\frac{3}{5}\qquad P(A')\approx\frac{32}{80}=\frac{2}{5}$$

有利于 A 的机会比率是

$$\frac{P(A)}{P(A')} \approx \frac{3/5}{2/5} = \frac{3}{2}，或近似 3 比 2$$

这是一个公平打赌，因为 \$30 比 \$20 恰好是 3 比 2.

补 充 习 题

概率的古典解释

8.29 有一个机会对策，掷一粒六面的骰子一次. 计算以下概率：(a)P(3 或 5)，(b)P(奇数点).

答案：(a)0.33，(b)0.50

8.30 有一个机会对策，从一副标准扑克牌中抽取一张. 计算以下概率：(a)P(Q)，(b)P(方块或红心)，(c)P(J，Q，或 K).

答案：(a)0.08，(b)0.50，(c)0.23

8.31 假设生日等可能出现在一年 365 天的任一天. 你刚结识一个新朋友. 则她出生在以下时间的概率是多少：(a)与你同一天出生，(b)与你同一月出生.

答案：(a)0.003，(b)0.08

8.32 20 个名字，包括你和你女友的，分别写在小纸片上并放入一顶帽子内. 将纸片完全混合后抽取一张. 从帽子中抽出以下名字的概率是多少：(a)你的名字，(b)你的或者你女友的名字.

答案：(a)0.05，(b)0.10

8.33 有 50 个人参加毕业舞会，包括 5 个祖父和 2 个祖母. 每个人的名字都分别写在小纸片上并放入一个盒子，完全混合. 闭上眼睛从盒子中抽取一个名字. 试问抽出以下名字的概率是多少：(a)一个祖父的名字，(b)一个祖母的名字.

答案：(a)0.10，(b)0.04

概率的相对频数解释

8.34 在 64 岁时，你了解到 620 个大学同学中已经有 40 个去世. 试问：在你就读的大学，今年的一个毕业生能够活到 64 岁的近似概率是多少？

答案：0.94

8.35 一家保险公司想了解汽车的挡风玻璃破碎而必须更换的概率. 公司收集了 20000 部汽车的信息，时间从一年的 7 月 1 日到下一年的 7 月 1 日，发现有 600 部汽车的挡风玻璃破碎. 在一年时间内，一部汽车的挡风玻璃破碎的近似概率是多少？

答案：0.03

8.36 一所大学想知道，在不超过秋季招生人数的情况下，能够接受的学生数. 该大学检查了过去 5 年的记录，发现在此期间共接受了 30000 名学生，但是其中仅有 12000 名在当年秋季注册入学. 该大学接受的一名学生将在秋季注册入学的概率是多少？

答案：0.40

8.37 一个菜农通常在 6 月 1 日种植西红柿，因为在这之后田地不会再发生霜冻. 今年，他计划在 5 月 15 日种植西红柿，因此想了解：在 5 月 15 日到 6 月 1 日之间出现一次霜冻的机会有多大？他从附近的气象台收集了过去 30 年的天气记录并进行检查，发现其中有 6 年在 5 月 15 日到 6 月 1 日之间出现过一次霜冻. 试问：在今年 5 月 15 日到 6 月 1 日之间，出现一次霜冻的近似概率是多少？

答案：0.20

8.38 一家制造计算机软盘的公司想了解生产出的软盘是不合格品的概率. 公司检查了 2000 张软盘，发现有 100 张不合格品. 试问：某张软盘是不合格品的近似概率？

答案：0.05

集合，子集，样本空间和事件

8.39 将 60 粒石子放入一只瓶子并完全混合在一起. 其中 20 粒红色，30 粒蓝色和 10 粒绿色. 试验是不进行观察取出一粒石子. 试问：试验的样本空间是什么？

答案：S={红色，蓝色，绿色}

8.40 一家乳制品公司通过抽奖促销产品，容器塞子下面的数字表明顾客的赢取奖金数.一些容器的塞子下面没有数字，而有些是＄1000、＄2000或＄5000.如果试验是从某个超市抽取该乳制品公司的一个容器，则样本空间是什么？

答案：S＝{没有数字，＄1000，＄2000，＄5000}

8.41 在掷硬币试验中，S＝{正面，反面}.试问：掷一次硬币能产生多少个可能事件？

答案：$2^n=2^2=4$

8.42 试验是掷骰子2次并以2次掷出的和作为结果.(a)描述包含事件A＝12的子集.它是简单事件还是复合事件？(b)描述包含事件A＝8的子集.它是简单事件还是复合事件？

答案：(a)A＝{6,6}，简单事件，(b)A＝{(1,7)，(2,6)，(3,5)，(4,4)}，复合事件

8.43 一所大学有5类学生：新生，二年级生，三年级生，四年级生和研究生.试验是随机选出3名学生，并以学生类别作为试验结果.试问：样本空间是什么，试验能产生多少个事件？

答案：S＝{新生，二年级生，三年级生，四年级生，研究生}；可能事件数＝$2^5=32$

8.44 一家葡萄酒酿造厂在以下四个年份：1986，1989，1991和1993年生产相同商标的葡萄酒，想知道顾客能否辨别不同年份的酒.为此，酒厂举行了一个品酒晚会，将四种不同的酒倒入200个相同的酒杯(每种酒50个)并且随机排放在桌上.试验是随意选出一个顾客请他选取4杯酒并且逐一品尝.试问：样本空间是什么，并且试验能产生多少个事件？

答案：S＝{1986，1989，1991，1993}；可能事件数＝$2^4=16$

Venn图

8.45 Venn图可以图示样本空间，样本空间中的事件，事件的补，事件的并和事件的交.定义集合的这些概念

答案：样本空间是试验的所有可能结果的集合.一个事件是试验的一个特定结果.一个事件的补是样本空间的如下子集：包含集合中不属于该事件的所有元素.事件A和B的并是集合的如下子集：其结果属于A或属于B或既属于A又属于B.事件A和B的交是集合的如下子集：其结果既属于A又属于B

8.46 一所大学提供30门历史方面的课程.集合是这些历史课程，从1到30.试验是选取两个毕业于历史专业的学生，并且列出他们学过哪些历史课程.事件A是学生A学过的一列课程，事件B是学生B学过的一列课程.用集合论符号表示以下试验结果：(a)学生A学过但学生B未学过的课程目录，(b)或者学生A学过或者学生B学过或者二者都学过的课程目录，(c)学生A和学生B都未学过的课程目录，(d)学生A和学生B都学过的课程目录.

答案：(a)$A\cup B'$，(b)$A\cup B$，(c)$A'\cup B'$，(d)$A\cap B$

8.47 两个朋友每年尽可能的多读一些新小说，并且两人从同一家书店买书.集合是去年出版并在该书店有售的所有新小说的一个目录.事件A是朋友A读过的新小说的目录，事件B是朋友B读过的新小说的目录.用集合论符号表示以下试验结果：(a)两个朋友都读过的新小说，(b)朋友A读过而朋友B未读过的新小说，(c)两个朋友都未读过的新小说.

答案：(a)$A\cap B$，(b)$A\cup B'$，(c)$A'\cup B'$

概率的集合论解释

8.48 试验是从一副完全洗乱的标准扑克牌中抽取一张，其中S＝{一副标准的52张扑克牌}，使用8.6节的公理和性质，计算抽出一张红色牌的概率.

答案：0.50

8.49 对习题8.48的抽牌试验，计算抽出一张红心或方块的概率.

答案：0.25＋0.25＝0.50

8.50 对习题8.48的抽牌试验，计算抽出一张红心和一张方块的概率.

答案：$P(\varnothing)=0$

8.51 对习题8.48的抽牌试验，计算抽出一张红心K的概率.

答案：0.019

概率的主观解释和机会比率的概念

8.52 一家保险公司估计某客户的住房明年发生一场火灾的概率近似为1/125.试问：这是一个古典，相对频

数或主观概率吗？使用来自 8.8 节的恰当公式将概率变换为机会比率.

答案：相对频数概率；明年该客户的住房不会发生一次火灾的机会比率是 124 比 1

8.53 掷一枚硬币 2 次，至少得到一次正面的概率是 0.75. 试问：这是一个古典，相对频数或主观概率吗？使用来自 8.8 节的恰当公式将概率变换为机会比率.

答案：古典概率；至少得到一次正面的机会比率是 3 比 1

8.54 你猜测一块有兴趣购买的地产，在随后的五年将有 40% 的机会地价至少上涨 60%. 试问：这是一个古典，相对频数或主观概率吗？使用来自 8.8 节的恰当公式将概率变换为机会比率.

答案：主观概率；在随后五年地价不会至少上涨 60% 的机会比率是 3 比 2

8.55 从一副标准扑克牌中抽取一张，取出牌面为 2 到 10 的一张牌的机会比率是 9 比 4. 使用来自 8.9 节的恰当公式将机会比率变换为概率，然后说明所得的概率是否为古典，相对频数，或主观概率.

答案：0.69；古典概率

8.56 据估计，钻井不能钻到石油的机会比率是 100 比 1. 使用来自 8.9 节的恰当公式将机会比率变换为概率，然后说明所得的概率是否为古典，相对频数，或主观概率.

答案：0.99；相对频数概率

8.57 你估计，从今天开始，一年后股市将高于目前水平的机会比率是 1 比 1. 使用来自 8.9 节的恰当公式将机会比率变换为概率，然后说明所得的概率是否为古典，相对频数，或主观概率.

答案：0.50；主观概率

第九章　计算法则和计数法则

9.1　事件组合的概率计算

第八章主要讲述概率的 4 种解释：古典概率，相对频数，集合论及主观解释. 其中，我们初步涉及如何计算事件的概率. 事件是样本空间的一个子集. 而且在 8.6 节我们给出了两互斥事件并的概率计算公式. 已知事件 A 和 B，由公理 III，或者称为特殊加法法则，有下式成立

$$P(A \cup B) = P(A) + P(B)$$

而且，对两个以上的互斥事件 $A_1, A_2, \cdots, A_k$，性质 4 给出下面的公式

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k)$$

特别地，对互斥事件 A 和它的逆事件 A'，性质 3 给出下面的公式

$$P(A) + P(A') = 1$$

我们还可以研究事件 A，B 的另一种组合形式，交($A \cap B$)，若事件 A，B 互斥，我们可以作最基本的概率计算[见题 8.20(h)]

$$P(A \cap B) = P(\phi) = 0$$

这一章，我们来计算并不互斥的多个事件交和并的概率. 为此，我们必须先讨论多个事件之间新的关系，即它们是**相关**或**独立**. 我们将给出它们的数学定义(见 9.4 节). 但现在我们可以定性的说，两事件独立是指一个事件的发生并不影响另一个事件发生的概率；如果两事件不独立，则它们是相关的. 为了得到他们的数学定义，完成这一章的其余工作，我们必须首先研究**条件概率**.

9.2　条件概率

已知事件 A 和 B，求它们交的概率 $P(A \cap B)$，也就是回答这样一个问题：事件 A 和 B 同时发生的概率. 另一方面，如果我们要求这些事件的**条件概率**，也就是要回答与此相关的另一个问题：已知事件 B 发生的条件下，事件 A 发生的概率. 或者相反的问题：已知事件 A 发生的条件下，事件 B 发生的概率.

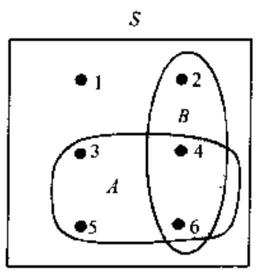


图 9-1

以掷骰子试验为例，说明条件概率不同于事件交的概率，样本空间为 $S=\{1,2,3,4,5,6\}$，$A=\{点数\geqslant 3\}$，$B=\{偶数\}$，用图 9-1 表示.

由 8.6 节性质 7，对任意一个包含 N 个等概率简单事件 e_i 的样本空间 S，有 $P(e_i)=\frac{1}{N}$和 $P(A)=N_A\left(\frac{1}{N}\right)$. 从而，对该样本空间有

$$P(e_i) = \frac{1}{N} = \frac{1}{6}$$

$$P(A) = N_A\left(\frac{1}{N}\right) = 4\left(\frac{1}{6}\right) = \frac{4}{6} = 0.67$$

$$P(B) = N_B\left(\frac{1}{N}\right) = 3\left(\frac{1}{6}\right) = \frac{3}{6} = 0.50$$

我们知道，事件 A 和 B 的交，$A \cap B$，是样本空间 S 的一个子集，由既属于 A 又属于 B 的元素组成[见例 8.5(b)]，用 $N_{A \cap B}$ 表示，则有

$$P(A \cap B) = P(A\text{ 和 }B) = N_{A\cap B}\left(\frac{1}{N}\right) = 2\left(\frac{1}{6}\right) = \frac{2}{6} = \frac{1}{3} = 0.33$$

在整个样本空间下，我们计算出三个事件的概率[$P(A)$，$P(B)$ 和 $P(A \cap B)$]；它们分别代

表感兴趣的事件所含样本点占样本空间 S 的样本点的比例.

现在,我们以同一个试验考虑条件概率.已知事件 B 发生的条件下,事件 A 发生的概率.在这个特定的掷骰子试验中,条件概率指已知出现偶数点(事件 B)的条件下,掷得的点数为 $\geqslant 3$(事件 A)的概率,用 $P(A|B)$ 表示这个条件概率,表达式中的竖线读作"已知"(或者"如果",见 8.3 节),整个表达式读作"已知 B 时,A 的概率".

条件概率 $P(A|B)$ 是 $A\cap B$ 中的样本点与 B 中的样本点的比.与计算 A 和 B 同时发生的概率不一样,它考虑的是整个样本空间 S,而在计算 B 发生的条件下 A 发生的概率时,只考虑样本空间的一部分一事件 B 所含的样本点,用数学符号表示为:

$$P(A \mid B)=\frac{N_{A\cap B}}{N_B} \tag{9.1}$$

这也可写成

$$P(A \mid B)=\frac{\dfrac{N_{A\cap B}}{N}}{\dfrac{N_B}{N}}$$

或者

$$P(A \mid B)=\frac{P(A\cap B)}{P(B)},\qquad 若\ P(B)\neq 0 \tag{9.2}$$

故后一个方程是**条件概率的一般形式**.它被用来计算已知事件 B 发生的条件下,事件 A 发生的概率.这个公式仅在 $P(B)\neq 0$ 的时候才适用,因为零作为除数是没有意义的.对上面的掷骰子试验

$$P(A \mid B)=\frac{N_{A\cap B}}{N_B}=\frac{2}{3}=0.67$$

也可写成

$$P(A \mid B)=\frac{P(A\cap B)}{P(B)}=\frac{1/3}{1/2}=\frac{2}{3}=0.67$$

0.67 这个概率是指已知掷得偶数点的条件下,所掷点数为 3,4,5,或 6 的概率.换句话说,三个偶数(2,4 和 6)中有两个是大于或等于 3 的.已知事件 B 发生的条件下事件 A 发生的概率(0.67)比事件 A 和 B 同时发生的概率(0.33)大.因为条件概率只考虑样本空间的一部分.

已知事件 B 发生的条件下事件 A 发生的概率,或者说 $P(A|B)$,是一个条件概率.事件 B 发生的概率 $P(B)$ 与事件 A 是否发生无关,因而它是一个**无条件概率**.也称为**简单概率**.

如果想用 $P(B|A)$ 代替 $P(A|B)$,则这个概率可写成

$$\begin{aligned}P(B \mid A)&=\frac{N_{A\cap B}}{N_A}\\&=\frac{P(A\cap B)}{P(A)},\qquad 若\ P(A)\neq 0\\&=\frac{1/3}{2/3}=\frac{1}{2}=0.50\end{aligned}$$

0.50 这个概率是指已知掷得的点数 $\geqslant 3$ 的条件下掷得偶数的概率.大于等于 3 的 4 个数中有 2 个是偶数.再次注意,该公式只有在 $P(A)\neq 0$ 时才成立.这里,$P(B|A)$ 是一个条件概率,而 $P(A)$ 是一个无条件概率.

例 9.1　一个医疗组开发了一种可能治疗普通感冒的疫苗.他们在 160 个志愿者身上测验.把这 160 个人分成两组,每组 80 人,其中一组为试验组,另一组为控制组.试验组的成员注射疫苗,而控制组的成员不注射.12 个月后,调查这 160 个人在去年是否患过感冒.结果归总在表 9.1(如,有 48

表 9.1

注射疫苗者	患感冒		总数
	是	否	
是	48	32	80
否	52	28	80
总数	100	60	160

位注射疫苗者患过感冒). 概率试验就是从这 160 人中随机地抽取一人,在该试验中,$S=\{160\text{ 个人}\}$,$A=\{\text{患过感冒}\}$,$B=\{\text{注射疫苗}\}$,求概率 $P(A|B)$.

解 用(9.2)式

$$P(A\mid B)=\frac{P(A\cap B)}{P(B)}$$

这里

$$N_B=80,\qquad N_{A\cap B}=48,\qquad \frac{1}{N}=\frac{1}{160}$$

我们求得

$$P(A\cap B)=N_{A\cap B}\left(\frac{1}{N}\right)=48\left(\frac{1}{160}\right)=\frac{48}{160}=\frac{3}{10}$$

$$P(B)=N_B\left(\frac{1}{N}\right)=80\left(\frac{1}{160}\right)=\frac{80}{160}=\frac{1}{2}$$

和

$$P(A\mid B)=\frac{P(A\cap B)}{P(B)}=\frac{3/10}{1/2}=\frac{3}{5}=0.60$$

这说明从注射疫苗组中随机地选取一人患有感冒的概率是 0.60.

例 9.2 对例 9.1 中的试验,$A'=\{\text{未患感冒}\}$和 $B'=\{\text{未注射疫苗}\}$. 用例 9.1 中的信息来计算 $P(A|B)+P(A'|B)$.

解 我们知道

$$N=60,P(B)=1/2,P(A\mid B)=0.60$$

而且由表 9.1,$N_{A'\cap B}=32$. 因此

$$P(A'\cap B)=N_{A'\cap B}\left(\frac{1}{N}\right)=32\left(\frac{1}{160}\right)=\frac{1}{5}$$

然后,由(9.2)式

$$P(A'\mid B)=\frac{P(A'\cap B)}{P(B)}=\frac{1/5}{1/2}=\frac{0.2}{0.5}=0.40$$

因此

$$P(A\mid B)+P(A'\mid B)=0.60+0.40=1.00$$

这说明注射疫苗者有 100%的概率患感冒或不患感冒.

注:这例子表明当一个事件和它的逆在同一个事件发生的条件下,下式一般是正确的:

$$P(Y\mid X)+P(Y'\mid X)=1$$

即无论条件事件是什么,一个事件发生或者它的逆事件发生这一现象以 100%的概率出现.

9.3 一般乘法法则

概率乘法法则是处理两个事件同时发生的概率计算问题的. 两个事件 A 和 B 同时发生的概率就是两事件交的概率,$P(A\cap B)$.

一般乘法法则用来计算相关事件交的概率(参见 9.1 节),从而也用来计算条件概率. 表达式 $P(A|B)$,表示已知事件 B 发生的条件下事件 A 发生的概率,可以看作事件 A 和 B 同时发生的次数与事件 B 发生的次数的比. 用概率符号即,

$$P(A\mid B)=\frac{P(A\cap B)}{P(B)}$$

或者,对于已知事件 A 发生的条件下事件 B 发生的概率

$$P(B\mid A)=\frac{P(A\cap B)}{P(A)}$$

变形得到求 $P(A\cap B)$的表达式

$$P(A\cap B)=P(B)P(A\mid B)\tag{9.3}$$

和

$$P(A\cap B)=P(A)P(B\mid A)\tag{9.4}$$

这两个数学表达式就是一般乘法法则. 简言之,这个公式表明,两个事件交的概率,等于两个事件中任一事件的无条件概率,和已知这一事件发生的条件下另一事件发生的条件概率的乘积.

例 9.3　在例 9.1 中,如果从 160 人中一个接一个地随机选出两个人,抽到的两人都注射过疫苗的概率是多少?

解☞　第二个人是注射疫苗者的概率依赖于第一个人是否为注射疫苗者,因为在挑选第二个人之前第一个人已经不在样本空间了. 因此,第二个事件的概率依赖于第一个事件的结果. 我们记 B_1 = {选出的第一个人是注射疫苗者}, B_2 = {选出的第二个人是注射疫苗者},用一般乘法法则得

$$P(B_1 \cap B_2) = P(B_1)P(B_2 \mid B_1)$$

由例 9.1 知

$$P(B_1) = P(B) = \frac{80}{160} = \frac{1}{2}$$

从注射疫苗者中去掉一个(见 3.16 节),得条件概率

$$P(B_2 \mid B_1) = \frac{79}{159}$$

和

$$\begin{aligned} P(B_1 \cap B_2) &= P(B_1)P(B_2 \mid B_1) \\ &= \frac{1}{2} \times \frac{79}{159} = \frac{79}{318} = 0.25 \end{aligned}$$

一般乘法法则可以推广到两个以上事件的情况. 这一推广被称作**一般乘法法则的推广**.

例 9.4　从一副扑克牌中任取三张,取后不放回,以任何次序,取得黑桃 J,红桃 Q,方块 K 三张的概率是多少?

解☞　我们要计算 k 个事件 $A_1, A_2, A_3, \cdots, A_k$ 交的概率,由一般乘法法则的推广得:

$$\begin{aligned} &P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_k) \\ &= P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_2 \cap A_1)\cdots P(A_k \mid A_{k-1} \cap \cdots \cap A_1) \end{aligned} \tag{9.5}$$

我们用 C_1 表示三张牌中第一张, C_2 为第二张, C_3 为第三张,那么

$$P(C_1) = \frac{3}{52}$$

$$P(C_2 \mid C_1) = \frac{2}{51}$$

$$P(C_3 \mid C_2 \cap C_1) = \frac{1}{50}$$

和

$$\begin{aligned} P(C_1 \cap C_2 \cap C_3) &= P(C_1)P(C_2 \mid C_1)P(C_3 \mid C_2 \cap C_1) \\ &= \frac{3}{52} \times \frac{2}{51} \times \frac{1}{50} = 0.000045 \end{aligned}$$

9.4　独立事件和相关事件

下面我们从数学上定义独立事件和相关事件. 回忆 9.1 节中,我们说两个事件**独立**是指一个事件的发生不影响另一个事件发生的**概率**. 这一叙述可以用下面的条件概率来表示,若事件 A, B 独立:

$$P(A \mid B) = P(A) \tag{9.6}$$

和

$$P(B \mid A) = P(B) \tag{9.7}$$

本质上说,已知事件 B 发生的条件下事件 A 发生的条件概率等于事件 A 发生的无条件概率. 如果这成立,那下面也成立:已知事件 A 发生的条件下事件 B 发生的条件概率等于事件 B 发生的无条件概率. 如果事件 A 与事件 B 独立,则事件 B 自然也与事件 A 独立,我们就称事件 A 和 B 是**独立事件**. 如果事件 A, B 不独立,则称事件 A, B 是**相关事件**.

9.5　特殊乘法法则

特殊乘法法则用于处理两个独立事件同时发生的概率. 从 9.4 节我们知道只要事件 A 和

B 是独立的,在(9.3)式就可以用 $P(A)$ 代替 $P(A|B)$ 得

$$P(A \cap B) = P(B)P(A \mid B) = P(B)P(A)$$

在(9.4)式中,用 $P(B)$ 代替 $P(B|A)$ 得

$$P(A \cap B) = P(A)P(B \mid A) = P(A)P(B)$$

这些等式称为**特殊乘法法则**,写为

$$P(A \cap B) = P(A)P(B) \tag{9.8}$$

例 9.5 用特殊乘法法则求下列概率:(a)在掷骰子试验中两次都掷得 4.(b)在重复抽取扑克牌的试验中,两次都抽取 Q 的概率,并假定每次抽取后均放回,且重新洗牌.

解☞

(a) $A=\{$第一次掷得 4$\}$, $P(A)=\frac{1}{6}$

$B=\{$第二次掷得 4$\}$, $P(B)=\frac{1}{6}$

$$P(A \cap B) = P(A)P(B) = \frac{1}{6} \times \frac{1}{6} = 0.028$$

(b) $A=\{$第一次取得 $Q\}$, $P(A)=\frac{4}{52}=\frac{1}{13}$

$B=\{$重新放回且洗牌后,第二次取得 $Q\}$, $P(B)=\frac{4}{52}=\frac{1}{13}$

$$P(A \cap B) = P(A)P(B) = \frac{1}{13} \times \frac{1}{13} = \frac{1}{169} = 0.0059$$

当求 k 个**独立事件** $A_1, A_2, A_3, \cdots, A_k$ 交的概率时,特殊乘法法则[见(9.8)式]推广为

$$P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_k) = P(A_1)P(A_2)P(A_3)\cdots P(A_k) \tag{9.9}$$

称为**特殊乘法法则的推广**.

9.6 一般加法法则

在 8.6 节的公理 III 中,我们给出两个互斥事件 A 和 B 并的特殊加法法则

$$P(A \cup B) = P(A) + P(B)$$

我们继续研究求事件 A 和 B 并的一般加法法则,但 A,B 不一定是互斥的.

由 8.6 节的性质 7 我们知道特殊加法法则可以写成

$$P(A \cup B) = N_A\left(\frac{1}{N}\right) + N_B\left(\frac{1}{N}\right)$$

这个等式只能用于计算两个互斥事件的并,如果用来计算两个相容事件的并时,事件 A 和 B 交中的样本点被计算了两次.故对于同一样本空间中的两相容事件(有共同的样本点),上面的式子修改如下

$$P(A \cup B) = N_A\left(\frac{1}{N}\right) + N_B\left(\frac{1}{N}\right) - N_B\left(\frac{1}{N}\right)$$

$$P(A \cup B) = N_{A \cup B}\left(\frac{1}{N}\right) = N_A + N_B - N_{A \cap B}$$

最后这个分式的分子是事件 A 中的样本点数加上事件 B 中的样本点数减去事件 $A \cap B$ 中的样本点数.所以,上式改写为

$$P(A \cup B) = \frac{N_A}{N} + \frac{N_B}{N} - \frac{N_{A \cap B}}{N}$$

因此

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{9.10}$$

这个公式就是**一般加法法则**.

例 9.6 在图 9-1 中显示的掷骰子试验中,求掷得结果是 $A=\{3,4,5,6\}$,或者 $B=\{2,4,6\}$ 的概率分别是多少?

解☞　这两个事件相交(4 和 6)，因此事件 A,B 不是互斥的

$$P(A)=N_A\left(\frac{1}{N}\right)=\frac{4}{6}$$

$$P(B)=N_B\left(\frac{1}{N}\right)=\frac{3}{6}$$

$$P(A\cap B)=N_{A\cap B}\left(\frac{1}{N}\right)=\frac{2}{6}$$

用一般加法法则有

$$\begin{aligned}P(A\cup B)&=P(A)+P(B)-P(A\cap B)\\&=\frac{4}{6}+\frac{3}{6}-\frac{2}{6}=\frac{4+3-2}{6}=\frac{5}{6}=0.83\end{aligned}$$

一般加法法则可以推广到两个以上事件的情况. 这个计算两个或两个以上事件并的概率的更一般的形式，称为**一般加法法则的推广**.

例 9.7　试验为每次从洗好的扑克牌中抽取一张，样本空间 $S=\{52\text{ 张}\}$，$A=\{\text{人头牌}\}$，$B=\{\text{黑色}\}$和 $C=\{\text{梅花}\}$，求 $P(A\cup B\cup C)$.

解☞　事件 A,B,C 不是互斥的，对这三个事件的并，利用一般加法法则的推广

$$\begin{aligned}P(A\cup B\cup C)=&P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)\\&-P(B\cap C)+P(A\cap B\cap C)\end{aligned}\tag{9.11}$$

对这个试验来说

$$A=\{\text{人头牌}\},\quad P(A)=\frac{12}{52}$$

$$B=\{\text{黑色}\},\quad P(B)=\frac{26}{52}$$

$$C=\{\text{梅花}\},\quad P(C)=\frac{13}{52}$$

$$A\cap B=\{\text{黑色人头牌}\},\quad P(A\cap B)=\frac{6}{52}$$

$$A\cap C=\{\text{梅花人头牌}\},\quad P(A\cap B)=\frac{3}{52}$$

$$B\cap C=\{\text{黑色梅花}\},\quad P(B\cap C)=\frac{13}{52}$$

$$A\cap B\cap C=\{\text{黑色梅花人头牌}\},\quad P(A\cap B\cap C)=\frac{3}{52}$$

因此

$$\begin{aligned}P(A\cup B\cup C)&=P(A)+P(B)+P(C)-P(A\cap B)\\&\quad-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)\\&=\frac{12}{52}+\frac{26}{52}+\frac{13}{52}-\frac{6}{52}-\frac{3}{52}-\frac{13}{52}+\frac{3}{52}\\&=\frac{12+26+13-6-3-13+3}{52}=\frac{32}{52}=0.62\end{aligned}$$

9.7　从一般加法法则中导出特殊加法法则

概率的**特殊加法法则**用于两个互斥事件的情况，即事件 A,B 的交为空集. 即

$$N_{A\cap B}=0$$

从而

$$P(A\cap B)=0$$

利用一般加法法则

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

$$P(A\cup B)=P(A)+P(B)-0$$

$$P(A\cup B)=P(A)+P(B)\tag{9.12}$$

因此，特殊加法法则是一般加法法则在事件互斥时的一种特殊情况.

例 9.8　试验为每次从洗好的扑克牌中抽取一张. 抽得一张 K 或一张 2,3,4 的概率是多少?

解☞　事件 $A=\{K\}$，事件 $B=\{2,3,4\}$

$$A\cap B=\{K\text{且为}2,3,4\}, A\cup B=\{KA;r2,3,4\}$$

$$P(A)=\frac{4}{52}$$

$$P(B)=\frac{3(4)}{52}=\frac{12}{52}$$

因为事件 A,B 互斥

$$P(A\cap B)=0$$

因此

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$
$$=\frac{4}{52}+\frac{12}{52}-0=\frac{16}{52}=0.31$$

9.8 列联表,联合概率表及边缘概率表

这一节我们处理下面一类试验,其所有可能结果都可以分成互斥且**完备**(它们的并为整个样本空间,见 2.4 节)的两个变量. 例如,试验为一个潜在消费者走进汽车销售处. 试验结果可以根据这个人的性别(男或女)和是否为买主进行分类. 做 100 次这样的实验,即 100 个人走进汽车销售处. 表 9.2 显示这 100 个结果的分类情况. 一看即知这 100 个人中有 70 个人是男性(M),男性中有 40 个人是买主(P),30 个人不是买主(P'). 这样一个概括的表称为**列联表**. 如果,两个变量之间具有一种独立—相关的关系(见 1.17 节),也说成一个变量是否**相依**(相关)于另一个变量. 在第 20 章将用这种表研究相依关系. 例如,在表 9.2 中我们要研究的相依关系是:汽车买主与性别相依吗?

表 9.2

目录	买主		总数
	是买主(P)	不是买主(P')	
男性(M)	40	30	70
女性(F)	10	20	30
总数	50	50	100

从这个列联表中看出,不难把试验中的频数值变为概率. 表 9.3 显示了这一变化. 用表 9.2 的每一个频数值除以 100(潜在顾客的人数)得到相对频率,这相对频数就可做为对应的可能事件的概率(参见 8.2 节). 这样得到的概率可能是近似值,但在本试验中,都是精确值. 因此,下一个潜在消费者有 0.4 的概率是男性买主,有 0.3 的概率是女性(可能是买主也可能不是),表格中间的每个单元是两个事件交的概率,因此,$P(P\cap M)=40/100=0.40$,$P(P'\cap F)=20/100=0.20$。两事件交的概率称为**联合概率**,因为它们是两事件联合发生的概率. 这就是为什么这样的表称作**联合概率表**的原因.

表 9.3

目录	买主		边缘概率
	是买主(P)	不是买主(P')	
男性(M)	0.40	0.30	0.70
女性(F)	0.10	0.20	0.30
边缘概率	0.50	0.50	1.00

表 9.3 的下边和右边的概率称为**边缘概率**,因为它们处于联合概率表中的边缘位置. 它们是在某单独的一行或一列的概率,与其他变量无关,因此它们是无条件概率或简单概率(参见 9.2 节). 从而,下一个潜在消费者是男性的概率为 $P(M)=0.70$,是买主的概率为 $P(P)=0.50$.

注意到表中的简单概率是对应的联合概率的和,因此,下一个进来的人是买主的概率是:$P(P)=P(P\cap M)+P(P\cap F)$. 推广这一结果得到**边缘概率公式**(译者注:也称全概率公式).

若事件 B 可能有互不相容且完备的方式 $A_i(i=1,2,\cdots,k)$ 发生,则事件 B 发生的边缘概率等于 k 个联合概率之和

$$P(B)=P(A_1\cap B)+P(A_2\cap B)+\cdots+P(A_k\cap B)$$

或者

$$P(B)=\sum_{i=1}^{k}P(A_i\cap B) \tag{9.13}$$

利用[(9.4)式]还可以写成

$$P(B)=\sum_{i=1}^{k}P(A_i)P(B\mid A_i) \tag{9.14}$$

利用联合概率表，用(9.2)式也可求条件概率. 在本例中，有

$$P(P\mid M)=\frac{P(P\cap M)}{P(M)}=\frac{0.40}{0.70}=0.57$$

这表明，在已知下一位潜在消费者为男性的条件下，它是买主的概率为 0.57.

例 9.9　电脑公司的某种微机元件有三个供应厂商(A_1,A_2,A_3)公司现有这种元件 20%来自 A_1，有 50%来自 A_2，有 30%来自 A_3. 知道每个供应厂商的不合格率：A_1 为 1%，A_2 为 0.5%，A_3 为 0.9%. 概率试验是从中随机地抽取一个，看它是否不合格. A_1,A_2,A_3 为选取一个来自该供应厂商的元件. 事件 B 为取得合格的，事件 B' 为取得不合格. 从这些信息构造一个联合概率表，包括以下联合概率：$P(A_1\cap B)$，$P(A_2\cap B)$，$P(A_3\cap B)$，$P(A_1\cap B')$，$P(A_2\cap B')$，$P(A_3\cap B')$；和以下边缘概率：$P(A_1)$，$P(A_2)$，$P(A_3)$，$P(B)$，$P(B')$.

解☞　我们把相对频率的估计作为精确的概率. 从而得到

$$P(A_1)=0.20,P(A_2)=0.50,P(A_3)=0.30$$

和

$$P(B'\mid A)=0.01,\quad P(B'\mid A_2)=0.005,\quad P(B'\mid A_3)=0.009$$

由这些在已知元件厂商的条件下，取得不合格的条件概率，我们可以计算已知元件厂商的条件下取得合格的条件概率. 因此

$$P(B\mid A_1)=1-P(B'\mid A_1)=1-0.010=0.990$$
$$P(B\mid A_2)=1-P(B'\mid A_2)=1-0.005=0.995$$
$$P(B\mid A_3)=1-P(B'\mid A_3)=1-0.009=0.991$$

利用(9.4)式，我们可以计算联合概率. 即

$$P(A_1\cap B)=P(A_1)P(B\mid A_1)=(0.20)(0.990)=0.1980$$
$$P(A_2\cap B)=P(A_2)P(B\mid A_2)=(0.50)(0.995)=0.4975$$
$$P(A_3\cap B)=P(A_3)P(B\mid A_3)=(0.30)(0.991)=0.2973$$

和

$$P(A_1\cap B')=P(A_1)P(B'\mid A_1)=(0.20)(0.01)=0.0020$$
$$P(A_2\cap B')=P(A_2)P(B'\mid A_2)=(0.50)(0.05)=0.0025$$
$$P(A_3\cap B')=P(A_3)P(B'\mid A_3)=(0.30)(0.009)=0.0027$$

利用(9.14)式，

$$\begin{aligned}P(B)&=\sum_{i=1}^{k}P(A_i)P(B\mid A_i)\\&=P(A_1)P(B\mid A_1)+P(A_2)P(B\mid A_2)+P(A_3)P(B\mid A_3)\\&=0.1980+0.4975+0.2973=0.9928\end{aligned}$$

和

$$\begin{aligned}P(B')&=\sum_{i=1}^{k}P(A_i)P(B'\mid A_i)\\&=P(A_1)P(B'\mid A_1)+P(A_2)P(B'\mid A_2)+P(A_3)P(B'\mid A_3)\\&=0.0020+0.0025+0.0027=0.0072\end{aligned}$$

现在我们已经得到所有要求的概率，从而可以做出一个完整的联合概率表(表 9.4).

表 9.4

不合格品	供应商			边缘概率
	(A_1)	(A_2)	(A_3)	
合格(B)	0.1980	0.4975	0.2973	0.9928
不合格(B')	0.0020	0.0025	0.0027	0.0072
边缘概率	0.20	0.50	0.30	1.00

9.9 BAYES 定理

在例 9.9 中，$P(B|A_1)$代表已知一个元件来自 A_1 厂商的条件下，它是合格产品的概率；$P(B'|A_1)$代表已知它来自 A_1 的条件下，不合格的概率. 从因果关系(见 1.19 节)，也就是求：已知该元件来自 A_1 提供的原因，结果是 B(或 B')的概率是多少？在这一节，我们将求下面这些概率：$P(A_1|B)$和 $P(A_1|B')$，与回答下列求因问题相反：已知结果 B(或 B')出现的条件下，元件来自 A_1(原因)的概率是多少？它们是两类相反的问题，因为前一种是由因索果，后一种是由果索因. 为解决后一种新的反问题，我们先导出 Bayes 定理，其中假设事件 B(或 B')以 k 个互斥且完备的方式 A_i 发生. 对事件 B 和 A_1，Bayes 定理是

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{\sum_{i=1}^{k} P(A_i)P(B \mid A_i)} \tag{9.15}$$

因为 Bayes 定理给出的是与原因相反的结果，所以它也被称为求**原因概率的** Bayes **定理**.

要导出上面的事件 B 和 A_1 的 Bayes 定理，我们从[(9.2)式]已知 B 时 A_1 的条件概率与已知 A_1 时 B 的条件概率得一般形式.

$$P(A_1 \mid B) = \frac{P(A_1 \cap B)}{P(B)}, \qquad \text{若 } P(B) \neq 0$$

$$P(B \mid A_1) = \frac{P(A_1 \cap B)}{P(A_1)}, \qquad \text{若 } P(A_1) \neq 0$$

得到 $P(A_1 \cap B)$的两个等式

$$P(A_1 \cap B) = P(B)P(A_1 \mid B)$$
$$P(A_1 \cap B) = P(A_1)P(B \mid A_1)$$

从而

$$P(B)P(A_1 \mid B) = P(A_1)P(B \mid A_1)$$

两边同除以 $P(B)$

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(B)}$$

最后用边缘概率公式[(9.14)式]代替分母中的 $P(B)$，就得到 Bayes 定理

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{\sum_{i=1}^{k} P(A_i)P(B \mid A_i)}$$

显然在已知条件下，定理中的 B 可换成 B'，或者是 A_k 中任何一个事件，例如对 A_2 来说，

$$P(A_2 \mid B) = \frac{P(A_2)P(B \mid A_2)}{\sum_{i=1}^{k} P(A_i)P(B \mid A_i)}$$

例 9.10　如果你作为一名统计员，在例 9.9 中的电脑公司工作，请你用 Bayes 定理计算某个不合格品来自供应商 A_1 的概率.

解☞　问题是求概率 $P(A_1|B')$，利用 Bayes 定理[(9.15)式]和例 9.9 的计算结果

$$P(A_1 \mid B') = \frac{P(A_1)P(B' \mid A_1)}{\sum_{i=1}^{k} P(A_i)P(B' \mid A_i)}$$

$$P(A_1 \mid B') = \frac{P(A_1)P(B' \mid A_1)}{P(A_1)P(B' \mid A_1) + P(A_2)P(B' \mid A_2) + P(A_3)P(B' \mid A_3)}$$
$$= \frac{0.0020}{0.0020 + 0.0025 + 0.0027} = \frac{0.0020}{0.0072} = 0.28$$

在 Bayes 定理的使用范围中，试验前已知的无条件概率称为**先验概率**. 在例 9.9 中，我们在试验前已先知道随机地抽取一个元件来自供应商 A_1 的概率是 $P(A_1) = 0.20$. 试验后，如果已知取得的是不合格品，来自 A_1 的概率为 $P(A_1|B') = 0.28$. 这称为**后验概率**，它是在知道试

验结果后，对 $P(A_1)$ 的修正. 注意在概率上的变化：预先已知 A_1 的概率 0.20（供应商 A_1 提供 20% 元件），但试验后，如果我们选取的是不合格品，它来自 A_1 的概率是 0.28.

在试验结束且取得新信息之后，对先验概率的修正是 Bayes 定理的基本用途. 这对主观概率（见 8.7 节）尤为重要，一个事件的先验概率是建立在个人相信程度基础之上的. 从而，一个商业总经理在根据已有的定量数据、直觉、判断和信念来分配财政收入决策结果的主观概率时，可以把 Bayes 定理作为一种证据积累来修正这个进一步决策的概率. 在商业统计领域，这个定理作为**贝叶斯决策分析**的基础，它用来处理在财政决策环境下，对总经理的主观概率的反复测试与修正.

9.10 树型图

树型图是一种计算一列事件概率的工具. 是乘法法则的直观表示形式，显示两个或两个以上事件交和交的概率（联合概率）公式. 在这个图中，每个事件的概率用一条线表示，称为一个**分支**. 这些分支的序列形成事件的交，被称为**路径**.

树型图 9-2(a) 显示了相依事件交的概率（条件概率）. 例如，从一副扑克中无放回地抽取两张卡的试验中，考虑事件 $A_1=\{红桃 K\}$，$A_2=\{不是红桃 K\}$，$B_1=\{红桃\}$，$B_2=\{不是红桃\}$. 求第一次抽取是红桃 K 且第二次抽取不是红桃的概率. 第一个分支表示第一个事件的无条件概率 $[P(A_1)]$，同一条路径的第二个分支表示第二个事件的条件概率 $[P(B_1|A_1)]$. 两个事件的交，两个事件同时发生的结果，即 $A_1\cap B_1$ 列在路径右边第一纵列，它的概率公式 [(9.4)式] 列在第二纵列，因此对事件 A_1、B_1 有

$$P(A_1\cap B_1)=P(A_1)P(B_1\mid A_1)$$

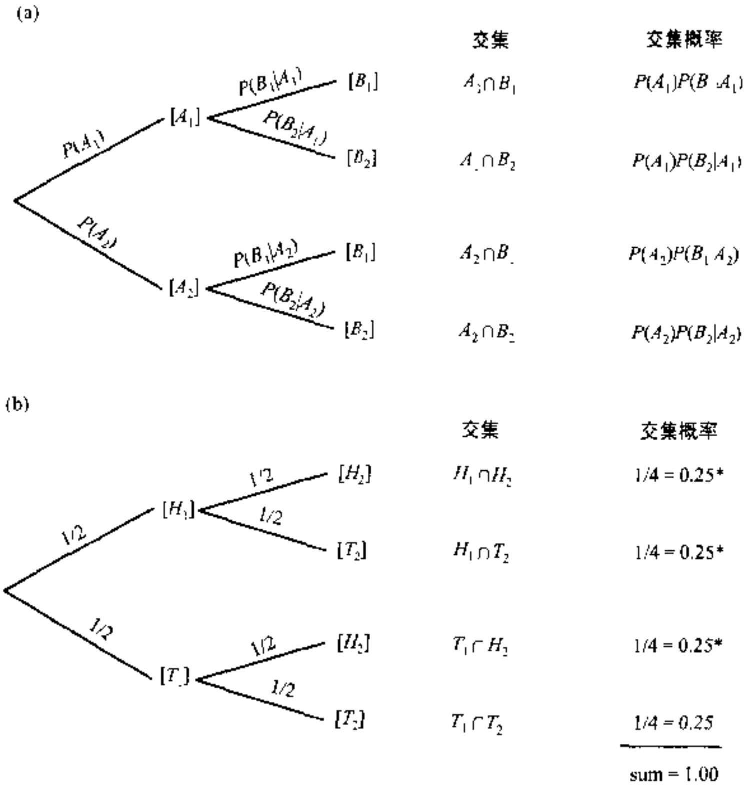


图 9-2

图 9-2(b) 是树型图中最典型的一种，将用在以后的问题中，它与图 9-2(a) 不同，在分支上显示的是概率值而不是概率符号，而且第二列给出的是事件交的概率值. 图 9-2(b) 显示的是掷两次硬币试验，由于第一次试验对第二次试验结果没有影响，所以两事件是独立的，它们交的概率可以用特殊乘法法则 [(9.8)式] 计算. 因此第一次和第二次都掷得正面的概率为

$$P(H_1 \cap H_2) = P(H_1)P(H_2) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$$

因为图中所列事件交包括了样本空间中所有互斥事件,因此它们的概率之和为 1.00.

例 9.11 试验为掷两次硬币,利用乘法法则和树型图,求两次掷得反面的概率以及两次不全掷得反面的概率.

解 单独一次掷得反面的概率为$\frac{1}{2}$,第一次掷得反面的事件(T_1)和第二次掷得反面的事件(T_2)是相互独立的.因此,应用特殊乘法法则得

$$P(T_1 \cap T_2) = P(T_1)P(T_2) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$$

图 9-2(b)中最下面一条路径表示 $P(T_1 \cap T_2)$的值与此相同.

要求两次掷得不全为反面的概率,通过先用特殊乘法法则得到除 $P(T_1 \cap T_2)$之外的所有事件交的概率来计算

$$P(H_1 \cap H_2) = P(H_1)P(H_2) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$$

$$P(H_1 \cap T_2) = P(H_1)P(T_2) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$$

$$P(T_1 \cap H_2) = P(T_1)P(H_2) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$$

然后利用 8.6 节性质 4 的变形,得到下面事件并的概率

$$P(H_1 \cap H_2) \cup P(H_1 \cap T_2) \cup P(T_1 \cap H_2)$$
$$= P(H_1 \cap H_2) + P(H_1 \cap T_2) + P(T_1 \cap H_2) = \frac{3}{4} = 0.75$$

根据树型图 9-2(b)所示,两次掷得不全为反面的概率是除了事件 $T_1 \cap T_2$ 之外所有交事件概率和,这些概率(在树型图中用 * 号标注)之和

$$P(H_1 \cap H_2) + P(H_1 \cap T_2) + P(T_1 \cap H_2) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} = 0.75$$

9.11 计数法则

根据 8.6 节性质 7,我们知道:如果样本空间 S 包括 N 个等概率简单事件(e_i),那么 $P(e_i) = \frac{1}{N}$和 $P(A) = N_A\left(\frac{1}{N}\right)$.然而样本空间的基本事件数(N)有时不易计算.**计数法则**是描述如何计数一个集合中事件的数学公式,在计算一列试验结果的概率时尤为重要.

计数法则有三种:**乘法原则**、**排列**和**组合**.对试验结果任意按唯一确定顺序放置就是结果的一种排列.当这种放置与顺序无关时就是一种组合.推导排列数和组合数的基本工具就是乘法原则.

9.12 计数法则:乘法原理

计数法则之一乘法原理,用来确定一个试验有两次或两次以上重复试验时的样本点总数.乘法原则最简单的一种形式可描述为:如果一个试验有两次连续的试验,第一次试验有 n_1 种可能结果;在第一次试验发生后,第二次试验有 n_2 种可能结果,那么样本点的总数(即两次联合试验的可能发生结果总数)为 $n_1 \times n_2$.把乘法原则推广到两个以上的试验中:如果试验有 k 次相继的试验,第一次有 n_1 种可能结果,第二次有 n_2 种可能结果,…,第 k 次有 n_k 种可能结果,用下面公式可以得到样本点总数

$$\# \text{ 样本点总数} = n_1 \times n_2 \times \cdots \times n_k \tag{9.16}$$

总之,一系列试验的可能结果相乘得到全体样本点总数.当一个样本空间用树型图来表示时,用乘法法则计算的是通过树型图所有单一路径的数目.

例 9.12 字母表中最后 4 个字母(W,X,Y,Z)能形成多少个三个字母的单词,每个字母可以重复使用,计算出结果,并用树型图检验其正确与否.

解 可以把该试验看作从 4 个字母中取 3 次.每次取出后又放回,下一次又可能取到同一个

字母，第一次试验有 4 种可能结果，(W, X, Y, Z)，第二次有同样的 4 种可能结果，第三次也是. 因此，$k=3$，$n_1=4$，$n_2=4$，$n_3=4$.

$$\# \text{样本点数} = n_1 \times n_2 \times n_3 = 4^3 = 64$$

它的树型图如图 9-3 所示，图中每一条路径都代表一个单词，和计算结果一样，这里有 64 条路径.

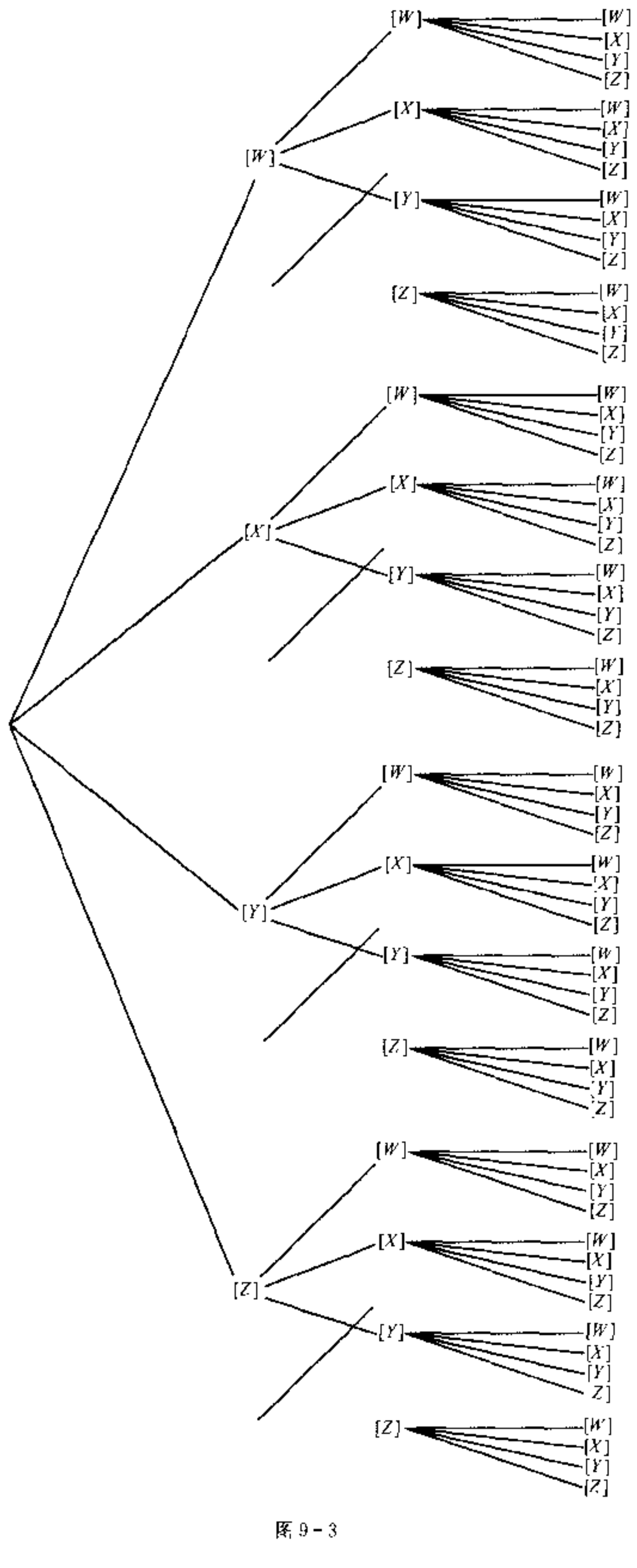


图 9-3

9.13 计数法则:排列

元素按一定的顺序放置物体就是元素的一种**排列**. **排列**是**计数法则**:乘法原则的一种特例. 它用来计算从一组元素中依次取出元素并按一定顺序排列的个数.

如果我们有 n 个不同的元素,顺次选取 r 个不同元素($r\leqslant n$),对第一次试验,我们可以取 n 个元素中任何一个. 对第二次试验,不能再选第一次选取的元素(因为它已经是该顺序中的第一个了),只能从剩下的 $n-1$ 个中任选一个. 对第三次试验,不能再选第一次和第二次选取的元素,因此,只能从剩下的 $n-2$ 个中任选一个如此进行下去,直到第 r 次试验,只能从剩下的 $n-r+1$ 个物体中任选一个. 所以,从 n 个物体中取出 r 个的所有不同顺序(排列)为

$$ {}_nP_r = n(n-1)\cdots(n-r+1) \tag{9.17} $$

这个关于${}_nP_r$ 的等式就是**计数法则中的排列公式**.(另一种符号:P_r^n 也指这个公式),可以用阶乘简化等式. 回忆 1.6 节中 n 的阶乘(写成 $n!$)计算形式

$$ n! = n(n-1)\cdots(n-r+1)(n-r)(n-r-1)\cdots(3)(2)(1) $$

因此

$$ \begin{aligned} n! &= [n(n-1)\cdots(n-r+1)][(n-r)(n-r-1)\cdots(3)(2)(1)] \\ &= {}_nP_r \times [(n-r)!] \end{aligned} $$

变形

$$ {}_nP_r = \frac{n!}{(n-r)!} \tag{9.18} $$

简化后的这个等式是计数法则的另一种形式:排列公式.

例 9.13 用(9.17)式和(9.18)式计算从字母表的最后 4 个字母(W,X,Y,Z)中取 3 个能组成多少个不同的单词? 同一个字母不能重复使用(不同于例 9.12). 并用树型图检验其结果的正确与否.

解答 问题是:从($n=4$)个字母中取($r=3$)个的排列是多少?

由(9.17)式

$$ {}_nP_r = n(n-1)\cdots(n-r+1) $$
$$ {}_4P_3 = 4(4-1)(4-3+1) = 4\times3\times2 = 24 $$

由(9.18)式

$$ {}_nP_r = \frac{n!}{(n-r)!} $$
$$ {}_4P_3 = \frac{4!}{(4-3)!} = \frac{4!}{1!} = \frac{4\times3\times2\times1}{1} = 24 $$

树型图(没有概率)如图 9-4 所示,树型图的每一条路径代表了从 4 个字母中取 3 个的一个排列,与计算结果一样有 24 种路径.

9.14 计数法则:组合

如果你有 n 个不同元素,并从中取出 r 个($r\leqslant n$),**不考虑顺序**,这样所得到的 r 个元素就是一个**组合**. 要计算从 n 个元素中一次取出 r 个不同元素的所有可能的组合种数. 我们需用**计数法则中的组合公式**. 它可以写作

$$ {}_nC_r = \frac{n(n-1)\cdots(n-r+1)}{r!} \tag{9.19} $$

或者

$$ {}_nC_r = \frac{n!}{r!(n-r)!} \tag{9.20} $$

(符号$\binom{n}{r}$和 C_r^n 也常用来表示这个公式.)

计数法则:组合是我们把树型图 9-4 中显示的 24 个三个字母的单词(即路径或者排列)

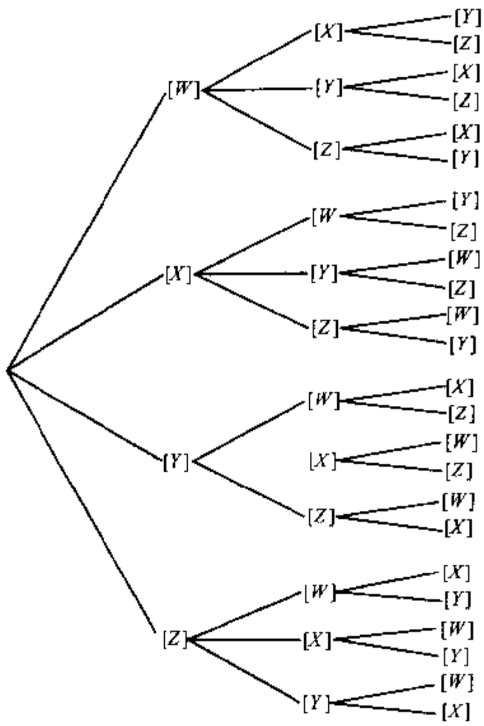


图 9-4

转换成图 9-5 的形式，分 4 行，每行有个 6 不同的单词，这样，每一行只有一种组合，却形成 6 种不同的排列. 由这个图看出，从 4 个字母中取 3 个不同字母的组合只有 4 种(每行一个).

例 9.14　利用(9.20)式求从字母表中的最后 4 个字母(W,X,Y,Z)中取 3 个的组合数

解

$$_nC_r=\frac{n!}{r!(n-r)!}=\frac{4!}{3!(4-3)!}=\frac{4\times3\times2\times1}{3\times2\times1(1)}=4$$

这一结果验证了图 9-5 中的结果.

WXY	WYX	XWY	XYW	YWX	YXW
WXZ	WZX	XWZ	XZW	ZWX	ZXW
WZX	WYZ	ZWY	ZYW	YWZ	YZW
XYZ	XZY	YXZ	YZX	ZXY	ZYX

图 9-5

计数法则:组合可看成计数法则:排列的一个特例. 利用计数法则:排列公式[(9.18)式]，我们可得到从 r 个元素($n=r$)取 r 个元素的排列种数是$\left[{}_rP_r=\frac{r!}{(r-r)!}=r!\right]$. 例如，3 个不同字母形成 3 个字母的单词数为($r!=3!=3\times2=6$). 由图 9-5 我们可以得到同样的结果，因为每一行包含由 3 个不同字母形成的所有 6 种排列，第一列包含从 4 个字母中每次取 3 个的 4 种组合，因此，我们可以看出排列的总数(${}_4P_3$)等于行数(${}_4C_3$)乘以列数($r!$).

$$_4P_3={}_4C_3(r!)=\left[\frac{4!}{3!(4-3)!}\right](3!)=\frac{4!}{1!}=24$$

这说明当从 n 个不同物体中取 r 个时，下式成立

$$_nP_r={}_nC_r\times r!$$

和

$$_nC_r=\frac{_nP_r}{r!}=\frac{n(n-1)\cdots(n-r+1)}{r!}=\frac{n!}{r!(n-r)!}$$

习 题 解 答

条件概率和乘法法则

9.1 考虑例 9.1 描述的试验,在 160 个志愿者身上测试一种疫苗.有 80 人注射疫苗,剩下 80 人未注射疫苗,12 个月之后,询问这 160 人在过去一年中是否患过感冒.其结果显示在表 9.1 中,试验就是从这 160 人中随机选取一人.该试验的样本空间 $S=\{160\text{人}\}$,$A=\{\text{患过感冒}\}$,$A'=\{\text{未患过感冒}\}$,$B=\{\text{注射疫苗者}\}$,$B'=\{\text{未注射疫苗者}\}$,求概率:(a)$P(A|B')$,(b)$P(B|A)$,(c)$P(B'|A')$.

解 ☞

(a)$P(A|B')=\dfrac{P(A\cap B)}{P(B')}$

$P(A\cap B')=N_{A\cap B'}\left(\dfrac{1}{N}\right)=52\left(\dfrac{1}{160}\right)=\dfrac{52}{160}=\dfrac{13}{40}$

$P(B')=N_{B'}\left(\dfrac{1}{N}\right)=80\left(\dfrac{1}{160}\right)=\dfrac{1}{2}$

从而

$$P(A\mid B')=\frac{P(A\cap B')}{P(B')}=\frac{13/40}{1/2}=\frac{13}{20}=0.65$$

这说明在未注射疫苗者中选取一人,患有感冒的概率为 0.65.

(b)$P(B|A)=\dfrac{P(A\cap B)}{P(A)}$

我们从例 9.1 知道

$$P(A\cap B)=\frac{3}{10}$$

由(8.1)式

$$P(A)=N_A\left(\frac{1}{N}\right)=100\left(\frac{1}{160}\right)=\frac{5}{8}$$

再由(9.2)式

$$P(B\mid A)=\frac{P(A\cap B)}{P(A)}=\frac{3/10}{5/8}=\frac{12}{25}=0.48$$

这说明在患感冒者中选取一人,注射疫苗的概率为 0.48.

(c)$P(B'|A')=\dfrac{P(A'\cap B')}{P(A')}$

$$P(A'\cap B')=N_{A'\cap B'}\left(\frac{1}{N}\right)=28\left(\frac{1}{160}\right)=\frac{7}{40}$$

$$P(A')=N_{A'}\left(\frac{1}{N}\right)=60\left(\frac{1}{160}\right)=\frac{3}{8}$$

从而

$$P(B'\mid A')=\frac{P(A'\cap B')}{P(A')}=\frac{7/40}{3/8}=\frac{7}{15}=0.47$$

这说明从未患感冒者中选取一人,未注射疫苗的概率是 0.47.

9.2 在习题 9.1 的试验中,随机地从 160 人中顺序取两人,求下面的概率:(a)二人都患感冒;(b)一人患感冒而另一人未患感冒

解 ☞

(a) 如果我们记事件 A_1 为第一人患感冒,A_2 事件为第二人患感冒,由(9.4)式得

$$P(A_1\cap A_2)=P(A_1)P(A_2\mid A_1)$$

由表 9.1 得

$$P(A_1)=N_{A_1}\left(\frac{1}{N}\right)=100\left(\frac{1}{160}\right)=\frac{100}{160}$$

已经从感冒组中选出一个人

$$P(A_2 \mid A_1) = \frac{99}{159}$$

从而

$$P(A_1 \cap A_2) = \frac{100}{160} \times \frac{99}{159} = \frac{9,900}{25,440} = 0.39$$

(b)我们记事件 A_1 为第一人患感冒，$A_2{}'$为第二人患感冒，由(9.4)式得

$$P(A_1 \cap A_2{}') = P(A_1)P(A_2{}' \mid A_1)$$

我们知道

$$P(A_1) = P(A) = \frac{100}{160}$$

仍有 60 人未患感冒，且从感冒组中选出 1 人

$$P(A_2{}' \mid A_1) = \frac{60}{159}$$

和

$$P(A_1 \cap A_2{}') = P(A_1)P(A_2{}' \mid A_1) = \frac{100}{160} \times \frac{60}{159} = \frac{6,000}{25,440} = 0.24$$

9.3 在一家大型工厂的雇员中，有 70%是高中毕业生，有 8%是管理人员，有 5%的既是管理员又是高中毕业生，根据以上信息，回答下列问题：(a)如果已知一个雇员是高中毕业生，那么他是管理人员的概率是多少？(b)如果已知这个雇员不是高中毕业生，那么他是管理人员的概率是多少？

解 ☞

(a)$A\cap B$＝{管理人员(A)且是高中毕业生(B)}，$P(A\cap B)=0.05$

B＝{高中毕业生}，$P(B)=0.70$

由(9.2)式

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.05}{0.70} = 0.071$$

(b)$A\cap B'$＝管理人员(A)且非高中毕业生(B')，$P(A\cap B')=0.03$

B'＝{非高中毕业生}，$P(B')=0.30$

$$P(A \mid B') = \frac{P(A \cap B')}{P(B')} = \frac{0.03}{0.30} = 0.10$$

9.4 连续掷两次骰子，求已知第一次掷得的点数是第二次的两倍的条件下，两次掷得的点数之和为 6 的概率是多少？

解 ☞　让事件 A＝{第一次掷得的点数是第二次的倍数}和 B＝{两次掷得的点数和为}．用(9.1)式

$$P(A \mid B) = \frac{N_{A\cap B}}{N_B}$$

这里

事件 $A\cap B$ 只有一种方式发生，第一次掷得 4，第二次掷得 2，因此 $N_{A\cap B}=1$

事件 B 有 5 种方式发生(1 和 5，5 和 1，2 和 4，4 和 2，3 和 3)，所以 $N_B=5$

因此

$$P(A \mid B) = \frac{1}{5} = 0.20$$

9.5 一副新洗好的标准纸牌，由方块、红桃、梅花和黑桃 4 组构成，每组有 13 张(从 A 到 K)，现不放回地从中抽取两张，求下面的概率：(a)两张都是红桃，(b)两张都是 Q，(c)一张是 Q 且另一张是 K？

解 ☞

(a)如果我们记事件 H_1 为第一次抽得红桃，事件 H_2 为第二次抽得红桃，然后利用(9.4)式

$$P(H_1 \cap H_2) = P(H_1)P(H_2 \mid H_1)$$

和

$$P(H_1) = \frac{13}{52} = \frac{1}{4}$$

$$P(H_2 \mid H_1) = \frac{12}{51}$$

因此

$$P(H_1 \cap H_2) = \frac{1}{4} \times \frac{12}{51} = \frac{12}{204} = 0.059$$

(b) 如果我们记事件 Q_1 为抽得的第一张是 Q,事件 Q_2 为第二张是 Q,利用(9.4)式,

$$P(Q_1 \cap Q_2) = P(Q_1)P(Q_2 \mid Q_1)$$

和

$$P(Q_1) = \frac{4}{52} = \frac{1}{13}$$

$$P(Q_2 \mid Q_1) = \frac{3}{51}$$

因此

$$P(Q_1 \cap Q_2) = \frac{1}{13} \times \frac{3}{51} = \frac{3}{663} = 0.0045$$

(c)如果我们记事件 K_1 为第一张是 K,事件 Q_2 为第二张是 Q,用(9.4)式

$$P(K_1 \cap Q_2) = P(K_1)P(Q_2 \mid K_1)$$

和

$$P(K_1) = \frac{4}{52} = \frac{1}{13}$$

$$P(Q_2 \mid K_1) = \frac{4}{51}$$

从而

$$P(K_1 \cap Q_2) = \frac{1}{13} \times \frac{4}{51} = \frac{4}{663} = 0.0060$$

9.6 自动头灯的一个制造商发给一个消费者 1000 台头灯,但他不知道其中有 3 台次品,这个消费者有一种检测方法,就是从这批货中随机地顺次抽取 3 台,如果这 3 台中至少有一台是次品,他就不接受这批货.求他拒绝这批货的概率是多少?

解☞ 我们感兴趣的事件是{3 台中至少有 1 台是次品},它的逆为{3 台中没有 1 台是次品},如果我们先用(9.5)式(即一般乘法法则的推广)得到逆事件的概率,然后用 8.6 节性质 3 就可以求得拒绝这批货的概率.

如果逆事件成立,则抽得的 3 台灯都是正品.用 N_1, N_2, N_3 分别表示这 3 次抽取头灯的事件.于是,我们发现

$$P(N_1) = \frac{997}{1{,}000}$$

$$P(N_2 \mid N_1) = \frac{996}{999}$$

$$P(N_3 \mid N_1 \cap N_2) = \frac{995}{998}$$

用(9.5)式

$$\begin{aligned} P(N_1 \cap N_2 \cap N_3) &= P(N_1)P(N_2 \mid N_1)P(N_3 \mid N_1 \cap N_2) \\ &= \frac{997}{1{,}000} \times \frac{996}{999} \times \frac{995}{998} = \frac{988{,}046{,}940}{997{,}002{,}000} = 0.991018 \end{aligned}$$

因此,利用 8.6 节性质 3,拒绝的概率是

$$\begin{aligned} P(\text{3 台中至少有一台是次品}) &= 1 - P(N_1 \cap N_2 \cap N_3) \\ &= 1 - 0.991018 = 0.0090 \end{aligned}$$

9.7 在第八章中,我们用概率的集合论解释来研究两次投币试验

$$P(\text{正正,正反,反正}) = P(\text{至少有一次正面}) = 0.75$$

现在,用(9.8)式(特殊乘法法则)来证明这种方法是正确的.

解☞ 如果我们记事件 H_1 和 H_2 代表第一次和第二次掷得正面,事件 T_1 和 T_2 代表第一次和第二次掷得反面,那么

$$P(\text{正正,正反,反正}) = P[(H_1 \cap H_2) \cup (H_1 \cap T_2) \cup (T_1 \cap H_2)]$$

又因为 8.6 节性质 4 有：如果 S 中 $A_1, A_2, \cdots, A_K$ 是互斥的，那么

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k)$$

从而

$$P(\text{正正,正反,反正}) = P(H_1 \cap H_2) + P(H_1 \cap T_2) + P(T_1 \cap H_2)$$

因为第一次掷得正面对第二次的结果没有影响，所以事件 H_1 和 H_2 是独立的从而利用(9.8)式得

$$P(H_1 \cap H_2) = P(H_1)P(H_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

同理，因为 H_1 和 T_2 是独立的

$$P(H_1 \cap T_2) = P(H_1)P(T_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

而且，因为事件 T_1 和 H_2 是独立的

$$P(T_1 \cap H_2) = P(T_1)P(H_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

从而

$$P(\text{正正,正反,正反}) = P(H_1 \cap H_2) + P(H_1 \cap T_2) + P(T_1 \cap H_2)$$
$$= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} = 0.75$$

9.8 如果已知 $P(A)=0.25$ 和 $P(A \cap B)=0.20$，在下列条件下分别计算 $P(B|A)$ 是多少？(a)事件 A 和 B 独立，(b)事件 A 和 B 不独立.

解 ☞

(a) A 和 B 独立，那么

$$P(B \mid A) = P(B)$$

用(9.8)式

$$P(A \cap B) = P(A)P(B)$$

从而

$$P(B \mid A) = P(B) = \frac{P(A \cap B)}{P(A)} = \frac{0.20}{0.25} = 0.80$$

(b)若 A 和 B 不独立，利用(9.4)式

$$P(A \cap B) = P(A)P(B \mid A)$$

和

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0.20}{0.25} = 0.80$$

9.9 在例 9.1 的感冒疫苗试验中，志愿者中有 100 人患有感冒，有 60 人未患感冒(参见表 9.1)，从这些人中随机地依次抽取 4 人，而且是有放回地，前两人患感冒且后两人未患感冒的概率是多少？

解 ☞　我们一般用(9.9)式(特殊乘法法则的推广)来计算 k 个独立事件的联合概率. 如果我们记事件 C_1, C_2 分别表示前两次独立抽取到患感冒的事件，事件 N_3 和 N_4 表示后两次独立抽取到未患感冒的事件，那么

$$P(C_1) = P(C_2) = \frac{100}{160} = \frac{5}{8}$$

和

$$P(N_3) = P(N_4) = \frac{60}{160} = \frac{3}{8}$$

从而

$$P(C_1 \cap C_2 \cap N_3 \cap N_4) = P(C_1)P(C_2)P(N_3)P(N_4)$$
$$= \frac{5}{8} \times \frac{5}{8} \times \frac{3}{8} \times \frac{3}{8} = \frac{225}{4{,}096} = 0.055$$

9.10 你掷 7 次硬币，7 次都掷得正面. (a)求这一事件发生的概率？(b)如果你继续掷第 8 次，掷得反面的概率是多少？

解 ☞ (a) 我们记 $H_1, H_2, H_3, H_4, H_5, H_6$ 和 H_7 分别为前7次独立掷得正面,那么

$$P(H_1)=P(H_2)=P(H_3)=P(H_4)=P(H_5)=P(H_6)=P(H_7)=\frac{1}{2}$$

由(9.9)式得

$$P(H_1\cap H_2\cap H_3\cap H_4\cap H_5\cap H_6\cap H_7)=P(H_1)P(H_2)P(H_3)P(H_4)P(H_5)P(H_6)P(H_7)$$
$$=\left(\frac{1}{2}\right)^7=\frac{1}{128}=0.0078$$

(b)每次单独投币的概率是相同的,因此,第8次掷得反面的概率是$\frac{1}{2}$,人们一般都错误的认为前7次都掷得正面,第8次掷得反面的概率比较大,这被称为赌徒的谬论.他的错误在于不接受各次投币都是独立的.

9.11 在依次掷3次骰子的试验中至少有一次掷得2的概率是多少?

解 ☞ 感兴趣的事件是{3次中至少有1次是2},它的逆为{3次没有1次掷得2},我们可以用(9.9)式求该逆事件的概率.根据8.6节性质3,我们就可以得到事件{3次中至少有一次掷得2}的概率,即对事件A和它的逆A',$P(A)+P(A')=1$.

如果该逆事件成立,即3次中每一次掷得的点数都不是2,用事件N_1, N_2, N_3分别表示,我们发现

$$P(N_1)=P(N_2)=P(N_3)=\frac{5}{6}$$

用(9.9)式

$$P(3\text{次中没有一次掷得}2)=P(N_1\cap N_2\cap N_3)=P(N_1)P(N_2)P(N_3)$$
$$=\left(\frac{5}{6}\right)^3=\frac{125}{216}=0.578704$$

从而,用性质3得,

$$P(3\text{次中至少有一次掷得}2)=1-P(3\text{次中没有一次掷得}2)$$
$$=1-0.578704\approx 0.42$$

加法法则

9.12 在例9.1的感冒疫苗试验中(参见表9.1),如果$S=\{160\text{人}\}$,$A=\{\text{患感冒}\}$,$A'=\{\text{未患感冒}\}$,$B=\{\text{注射疫苗}\}$,$B'=\{\text{未注射疫苗}\}$,用(9.10)式(一般加法法则)来求:(a)$P(A\cup B)$,(b)$P(A'\cup B)$,(c)$P(A\cup B')$,(d)$P(A'\cup B')$,(e)$P(A\cup A')$.

解 ☞

(a)$A=\{\text{患感冒}\}$, $P(A)=\frac{100}{160}$

$B=\{\text{注射疫苗}\}$, $P(B)=\frac{80}{160}$

$A\cap B=\{\text{患感冒而且是注射疫苗}\}$, $p(A\cap B)=\frac{48}{160}$

从而用(9.10)式得,

$$P(A\cap B)=P(A)+P(B)-P(A\cap B)=\frac{100}{160}+\frac{80}{160}-\frac{48}{160}=\frac{132}{160}=0.82$$

(b)$A'=\{\text{未患感冒}\}$, $P(A')=\frac{60}{160}$

$B=\{\text{注射疫苗}\}$, $P(B)=\frac{80}{160}$

$A'\cap B=\{\text{未患感冒但是注射疫苗}\}$, $P(A'\cap B)=\frac{32}{160}$

因此

$$P(A'\cup B)=\frac{60}{160}+\frac{80}{160}-\frac{32}{160}=\frac{108}{160}=0.68$$

(c)$A'=\{\text{患感冒}\}$, $P(A)=\frac{100}{160}$

$B'=\{\text{未注射疫苗}\}$, $P(B')=\frac{80}{160}$

$A\cap B'$={患感冒但未注射疫苗}，$P(A\cap B')=\frac{52}{160}$

因此

$$P(A\cup B')=\frac{100}{160}+\frac{80}{160}-\frac{52}{160}=\frac{128}{160}=0.80$$

(d)A'={未患感冒}，$P(A')=\frac{60}{160}$

B'={未注射疫苗}，$P(B')=\frac{80}{160}$

$A'\cap B'$={未患感冒而且也未注射疫苗}，$P(A'\cap B')=\frac{28}{160}$

因此

$$P(A'\cup B')=\frac{60}{160}+\frac{80}{160}-\frac{28}{160}=\frac{112}{160}=0.70$$

(e)A={患感冒}，$P(A)=\frac{100}{160}$

A'={未患感冒}，$P(A')=\frac{60}{160}$

$A\cap A'$={患感冒而且未患感冒}，$P(A\cap A')=0$，因为它们是互斥的

从而

$$P(A\cup A')=\frac{100}{160}+\frac{60}{160}-0=\frac{160}{160}=1.0$$

9.13 在图 9－6 中，圆圈上的数表示该圆所代表事件的概率，圆内的数为该面积代表的事件的概率，用这些信息和(9.10)式求：(a)$P(A\cup B)$，(b)$P(A'\cup B)$，(c)$P(A\cup B')$，(d)$P(A'\cup B')$.

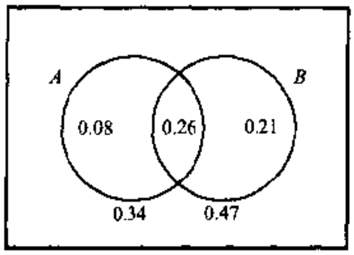


图 9－6

解

(a)$P(A)=0.34$，$P(B)=0.47$，$P(A\cap B)=0.26$

因此，用(9.10)式

$$\begin{aligned}P(A\cup B)&=P(A)+P(B)-P(A\cap B)\\&=0.34+0.47-0.26=0.55\end{aligned}$$

(b)$P(A')=1-0.34=0.66$，$P(B)=0.47$，$P(A'\cap B)=0.21$

因此

$$P(A'\cup B)=0.66+0.47-0.21=0.92$$

(c)$P(A)=0.34$，$P(B')=1-0.47=0.53$，$P(A\cap B')=0.08$

因此

$$P(A\cup B')=0.34+0.53-0.08=0.79$$

(d)$P(A')=0.66$，$P(B')=0.53$，$P(A'\cap B')=1-(0.08+0.26+0.21)=0.45$

因此

$$P(A'\cup B')=0.66+0.53-0.45=0.74$$

9.14 在图 9－6 显示的样本空间中，事件 A 和 B 是独立的吗？事件 A 和 B 是互斥的吗？

解 特殊乘法法则[(9.8)式]可以用来检验两个事件是否独立．如果两事件独立，则有下式成立

$$P(A\cap B)=P(A)P(B)$$

现在 $P(A)=0.34$，$P(B)=0.47$，因此 $P(A)P(B)=(0.34)(0.47)=0.1598$．但我们从上图知道

$$P(A\cap B)=0.26$$

因此，$P(A\cap B)\neq P(A)P(B)$，这两个事件是相关的．又因为 $P(A\cap B)\neq 0$，所以事件不是互斥的．

9.15 在图 9－7 的 Venn 图中，每个圆内的数表示该圆代表的事件的概率，利用这些已知条件，选择合适的加法法则，求：(a)$P(A\cup B)$，(b)$P(A'\cup B)$．

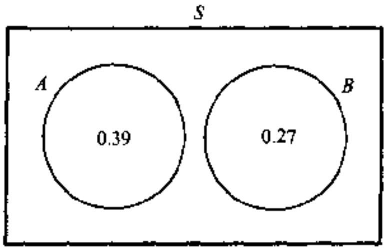


图 9－7

(c)$P(A\cup B')$,(d)$P(A'\cup B')$.

解

(a)$P(A)=0.39, P(B)=0.27$

因为这些事件是互斥的,我们可以利用(9.12)式

$$P(A\cup B)=P(A)+P(B)=0.39+0.27=0.66$$

(b)$P(A')=1-0.39=0.61, P(B)=0.27, P(A'\cap B)=0.27$

因为事件A',B不是互斥的,我们利用(9.10)式

$$P(A'\cup B)=P(A')+P(B)-P(A'\cap B)=0.61+0.27-0.27=0.61$$

(c)$P(A)=0.39, P(B')=1-0.27=0.73, P(A\cap B')=0.39$

又因为事件A,B'不是互斥的,我们用(9.10)式

$$P(A\cup B')=0.39+0.73-0.39=0.73$$

(d)$P(A')=0.71, P(B')=0.73, P(A'\cap B')=1-(0.39+0.27)=0.34$

再因为事件A',B'不是互斥的,我们利用(9.10)式

$$P(A'\cup B')=0.61+0.73-0.34=1.0$$

9.16　对图9-7显示的样本空间,事件A,B是独立的吗?

解　用(9.8)式来检验其独立性

$$P(A\cap B)=P(A)P(B)$$

在这个样本空间中,$P(A)=0.39, P(B)=0.27$,因此

$$P(A)P(B)=(0.39)(0.27)=0.1053$$

但从图我们知道

$$P(A\cap B)=0$$

因此,$P(A\cap B)\neq P(A)P(B)$,因而这两个事件是相关的.

9.17　在掷骰子的试验中,如果$S=\{1,2,3,4,5,6\}$,$A=\{$掷得偶数$\}$,$B=\{$掷得点数$\geqslant 2\}$,$C=\{$掷得点数$\leqslant 4\}$,利用(9.11)式求$P(A\cup B\cup C)$.

解　我们由已知条件知道

$$A=\{\text{偶数}\},\quad P(A)=\frac{3}{6}$$

$$B=\{\text{点数}\geqslant 2\},\quad P(B)=\frac{5}{6}$$

$$C=\{\text{点数}\leqslant 4\},\quad P(C)=\frac{4}{6}$$

$$A\cap B=\{\text{偶数}\geqslant 2\},\quad P(A\cap B)=\frac{3}{6}$$

$$A\cap C=\{\text{偶数}\leqslant 4\},\quad P(A\cap C)=\frac{2}{6}$$

$$B\cap C=\{2\leqslant\text{点数}\leqslant 4\},\quad P(B\cap C)=\frac{3}{6}$$

$$A\cap B\cap C=\{2\leqslant\text{偶数}\leqslant 4\},\quad P(A\cap B\cap C)=\frac{2}{6}$$

因此,用(9.11)式,

$$\begin{aligned}P(A\cup B\cup C)&=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)\\&\quad-P(B\cap C)+P(A\cap B\cap C)\\&=\frac{3}{6}+\frac{5}{6}+\frac{4}{6}-\frac{3}{6}-\frac{2}{6}-\frac{3}{6}+\frac{2}{6}\\&=\frac{3+5+4-3-2-3+2}{6}=\frac{6}{6}=1.0\end{aligned}$$

联合概率表,边缘概率和BAYES定理

9.18　有一种不透明的罐,看不见它里面的东西,现有3个这样的罐(A_1,A_2,A_3),每个罐中有4个球,这些球的颜色不一样,A_1中有3个蓝球1个黄球,A_2中有2个蓝球和2个黄

球，A_3 中有 1 个蓝球和 3 个黄球. 试验是随机地选一个罐，然后再从中任取一球，记 A_1, A_2, A_3 为分别从相应的罐中任取一个球的事件. B 表示取得篮球，B' 表示取得黄球. 由这些已知条件，做出联合概率表.

解 我们从已知条件得

$$P(A_1)=P(A_2)=P(A_3)=\frac{1}{3}=0.3333$$
$$P(B\mid A_1)=\frac{3}{4}=0.75$$
$$P(B\mid A_2)=\frac{2}{4}=0.50$$
$$P(B\mid A_3)=\frac{1}{4}=0.25$$
$$P(B'\mid A_1)=\frac{1}{4}=0.25$$
$$P(B'\mid A_2)=\frac{2}{4}=0.50$$
$$P(B'\mid A_3)=\frac{3}{4}=0.75$$

现在，我们用(9.3)式来计算联合概率

$$P(B\cap A_1)=P(A_1)P(B\mid A_1)=(0.3333)(0.75)=0.249975$$
$$P(B\cap A_2)=P(A_2)P(B\mid A_2)=(0.3333)(0.50)=0.166650$$
$$P(B\cap A_3)=P(A_3)P(B\mid A_3)=(0.3333)(0.25)=0.083325$$
$$P(B'\cap A_1)=P(A_1)P(B'\mid A_1)=(0.3333)(0.25)=0.083325$$
$$P(B'\cap A_2)=P(A_2)P(B'\mid A_2)=(0.3333)(0.50)=0.166650$$
$$P(B'\cap A_3)=P(A_3)P(B'\mid A_3)=(0.3333)(0.75)=0.249975$$

我们已知边缘概率 $P(A_1)$，$P(A_2)$和 $P(A_3)$，用(9.14)式(边缘概率公式)来计算 $P(B)$和$P(B')$. 因此

$$\begin{aligned}P(B)&=\sum_{i=1}^{k}P(A_i)P(B\mid A_i)\\&=P(A_1)P(B\mid A_1)+P(A_2)P(B\mid A_2)+P(A_3)P(B\mid A_3)\\&=0.249975+0.166650+0.083325=0.499950\end{aligned}$$

$$\begin{aligned}P(B')&=\sum_{i=1}^{k}P(A_i)P(B'\mid A_i)\\&=P(A_1)P(B'\mid A_1)+P(A_2)P(B'\mid A_2)+P(A_3)P(B'\mid A_3)\\&=0.083325+0.166650+249975=0.499950\end{aligned}$$

我们已得到所要求的概率，可以做出整个联合概率表，见表 9.5.

表 9.5

球	罐			边缘概率
	A_1	A_2	A_3	
B	0.249975	0.166650	0.083325	0.499950
B'	0.083325	0.166650	0.249975	0.499750
边缘概率	0.3333	0.3333	0.3333	1.00

9.19 根据习题 9.18 所提供的信息，直接由表 9.5 或者(9.15)式(Bayes 定理)都可以直接回答下列问题. (a)已知选取的是一个蓝球，求它取自罐 A_1 的概率是多少？(b)已知选取的是一黄球，求它是取自罐 A_2 的概率是多少？

解

(a)问题：$P(A_1\mid B)$是多少？

我们由(9.2)式知

$$P(A_1\mid B)=\frac{P(B\cap A_1)}{P(B)}$$

因此，根据表 9.5 得，

$$P(A_1 \mid B)=\frac{0.249975}{0.499950}=0.50$$

用(9.15)式和习题 9.18 的信息得,

$$P(A_1 \mid B)=\frac{P(A_1)P(B \mid A_1)}{\sum_{i=1}^{k}P(A_i)P(B \mid A_i)}$$

$$=\frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1)+P(A_2)P(B \mid A_2)+P(A_3)P(B \mid A_3)}$$

$$=\frac{0.249975}{0.249975+0.166650+0.083325}=\frac{0.249975}{0.499950}=0.50$$

这些结果表明:选中 A_1 的先验概率为 0.3333,一旦已知选中的是蓝球,选中 A_1 的概率上升为 0.50.

(b)问题:$P(A_2|B')$是多少?

由(9.2)式知道

$$P(A_2 \mid B')=\frac{P(B' \cap A_2)}{P(B')}$$

从而由表 9.5

$$P(A_2 \mid B')=\frac{0.166650}{0.499950}=0.3333$$

用(9.15)式和习题 9.18 的信息得,

$$P(A_2 \mid B')=\frac{P(A_2)P(B' \mid A_2)}{P(A_1)P(B' \mid A_1)+P(A_2)P(B' \mid A_2)+P(A_3)P(B' \mid A_3)}$$

$$=\frac{0.166650}{0.083325+0.166650+0.249975}=\frac{0.166650}{0.499950}=0.3333$$

这些结果表明选中 A_2 罐的先验概率和后验概率均为 0.3333.

9.20 某教材出版公司的一地区销售经理认为他的一家竞争公司有 60%的概率把他的化学教材出售给一所大学的化学系.同时他认为如果这一事件发生,则该大学所在城市的社区学院有 80%的概率也采用这种教材.如果这所大学不接受,社区学院还有 50%的概率采用种教材.如果令 U,U'分别代表大学采用和不用这家竞争公司的教材,而 C,C'为学院采用和不用这家竞争公司的教材,下面我们就可以做出一个联合概率表,且包括以下事件交的概率:$P(C\cap U)$,$P(C\cap U')$,$P(C'\cap U)$,$P(C'\cap U')$;和边缘概率:$P(U)$,$P(U')$,$P(C)$,$P(C')$.

解 ☞　从已知条件我们知道

$$P(U)=0.60,\qquad P(C \mid U)=0.80,\qquad P(C \mid U')=0.50$$

因此可以计算

$$P(U')=1-0.60=0.40$$
$$P(C' \mid U)=1-0.80=0.20$$
$$P(C' \mid U')=1-0.50=0.50$$

现在可以用(9.3)式计算事件交的概率

$$P(C \cap U)=P(U)P(C \mid U)=(0.60)(0.80)=0.48$$
$$P(C \cap U')=P(U')P(C \mid U')=(0.40)(0.50)=0.20$$
$$P(C' \cap U)=P(U)P(C' \mid U)=(0.60)(0.20)=0.12$$
$$P(C' \cap U')=P(U')P(C' \mid U')=(0.40)(0.50)=0.20$$

又已知边缘概率 $P(U)$,$P(U')$,那么利用(9.14)式就可以计算 $P(C)$,$P(C')$

$$P(C)=\sum_{i=1}^{k}P(U_i)P(C \mid U_i)=P(U)P(C \mid U)+P(U')P(C \mid U')$$
$$=0.48+0.20=0.68$$

$$P(C')=\sum_{i=1}^{k}P(U_i)P(C' \mid U_i)=P(U)P(C' \mid U)+P(U')P(C' \mid U')$$
$$=0.12+0.20=0.32$$

我们现在已经知道所有要求的概率,就可以做出完整的联合概率表,见表 9.6.

表 9.6

学院	大学		边缘概率
	U	U'	
C	0.48	0.20	0.68
C'	0.12	0.20	0.32
边缘概率	0.60	0.40	1.00

9.21 根据习题 9.20 的已知条件，由表 9.6 和(9.15)式，直接回答下列问题.(a)已知这所学院采用这种教材的条件下，这所大学也采用的概率是多少?(b)已知这所学院不用这种教材的条件下，这所大学也不用的概率是多少?

解 ☞

(a)问题:$P(U|C)$是多少?

从(9.2)式我们知道

$$P(U \mid C) = \frac{P(C \cap U)}{P(C)}$$

从而，由表 9.6 得

$$P(U \mid C) = \frac{0.48}{0.68} = 0.71$$

利用(9.15)式和习题 9.20 的已知条件

$$P(U \mid C) = \frac{P(U)P(C \mid U)}{\sum_{i=1}^{k} P(U_i)P(C \mid U_i)} = \frac{P(U)P(C \mid U)}{P(U)P(C \mid U) + P(U')P(C \mid U')}$$

$$= \frac{0.48}{0.48 + 0.20} = 0.71$$

以上结果表明该大学采用这种教材的先验概率为 0.60，在该社区学院采用这种教材的条件下，该 大学采用此种教材的后验概率上升为 0.71.

(b)问题:$P(U'|C')$是多少?

从(9.2)式知

$$P(U' \mid C') = \frac{P(C' \cap U')}{P(C')}$$

由表 9.6 得

$$P(U' \mid C') = \frac{0.20}{0.32} = 0.625,\text{或 } 0.62$$

利用(9.15)式和习题 9.20 的已知条件，

$$P(U' \mid C') = \frac{P(U')P(C' \mid U')}{P(U')P(C' \mid U') + P(U)P(C' \mid U)}$$

$$= \frac{0.20}{0.20 + 0.12} = \frac{0.20}{0.32} = 0.625,\text{或 } 0.62$$

以上结果表明该大学不用此种教材的先验(主观)概率为 0.40，在该社区学院不用这种教材的条件下，该大学拒绝的后验概率上升为 0.62.

9.22 一种病毒性疾病感染了南部 12 个州农场近 25%的猪群，现在对该病毒进行诊断测试，当一头猪确实患有该疾病时，诊断得到肯定结果(被病毒感染)的概率为 84%. 当一头猪未患这种疾病时，诊断得到否定结果(未被病毒感染)的概率为 80%. 概率试验就是从猪群中随机地挑选一头猪，测试是否感染病毒，根据这些已知条件，回答下列问题:(a) 如果测试结果是肯定的，则这头猪确实被病毒感染的概率有多大?(b) 如果测试结果是否定的，则这头猪未被病毒感染的概率有多大?(c)测试得到的结果是正确的概率有多大?

解 ☞　要回答这些问题，用 V 代表这头猪被病毒感染，V' 代表未被病毒感染，R 为测试得到肯定结果，R' 为测试得到否定结果.

(a)问题:$P(V|R)$是多少?

利用(9.15)式

$$P(V|R)=\frac{P(V)P(R|V)}{P(V)P(R|V)+P(V')P(R|V')}$$

虽然已知的百分数为相对频率的估计,但我们把它视作概率的准确值.因此,从已知条件可得

$$P(V)=0.25, P(R|V)=0.84, P(R'|V')=0.80$$

因此可以计算

$$P(V')=1-P(V)=1-0.25=0.75$$

$$P(R|V')=1-P(R'|V')=1-0.80=0.20$$

从而

$$P(V|R)=\frac{(0.25)(0.84)}{(0.25)(0.84)+(0.75)(0.20)}=\frac{0.21}{0.21+0.15}=\frac{0.21}{0.36}=0.58$$

所以,如果这头猪被感染的话,诊断得到肯定结果的概率为84%,反过来,当诊断得到肯定结果的话,这头猪确实被感染的概率只有58%.

(b)问题:$P(V'|R')$是多少?

利用(9.15)式

$$P(V'|R')=\frac{P(V')P(R'|V')}{P(V')P(R'|V')+P(V)P(R'|V)}$$

唯一剩下的是

$$P(R'|V)=1-P(R|V)=1-0.84=0.16$$

因此

$$P(V'|R')=\frac{(0.75)(0.80)}{(0.75)(0.80)+(0.25)(0.16)}=\frac{0.60}{0.60+0.04}=\frac{0.60}{0.64}=0.94$$

所以,如果这头猪确实未被感染的话,诊断得到否定结果的概率为80%;反过来,当诊断得到否定结果的话,这头猪确实未被感染的概率为94%.

(c)问题:$P[(V\cap R)\cup(V'\cap R')]$是多少?

因为$V\cap R$和$V'\cap R'$是互斥的,由8.6节性质4知道

$$P[(V\cap R)\cup(V'\cap R')]=P(V\cap R)+P(V'\cap R')$$

而且根据(9.4)式

$$P(V\cap R)=P(V)P(R|V)=(0.25)(0.84)=0.21$$

$$P(V'\cap R')=P(V')P(R'|V')=(0.75)(0.80)=0.60$$

因此

$$P(\text{正确诊断})=P[(V\cap R)\cup(V'\cap R')]=0.21+0.60=0.81$$

因此,该测试方法有81%的概率得到正确结果。

9.23 一保险公司的总经理发明一种对保险能力的测试方法,她知道,现有的销售人员中有65%有良好销售记录,有35%有不好记录,他让所有的销售人员都做这项测试,发现有良好记录的销售人员中有73%通过该测试,而那些有不好记录的销售人员中有78%的未通过这项测试。现从中随机的抽取一人,让他做这项测试.根据以上信息回答下列问题.(a)如果此人通过这项测试,则他有良好销售记录的概率是多少?(b)如此人未通过测试,则他有不好销售记录的概率是多少?(c)测试结果和他原有销售记录一致的概率是多少?

解 ☞ 用T和T'分别表示是否通过测试,R和R'表示具有良好销售记录或不好销售记录.

(a)问题:$P(R|T)$是多少?

利用(9.15)式,

$$P(R|T)=\frac{P(R)P(T|R)}{P(R)P(T|R)+P(R')P(T|R')}$$

由已知条件有

$$P(R)=0.65, P(R')=0.35, \qquad P(T|R)=0.73, P(T'|R')=0.78$$

因此我们可以计算

$$P(T|R')=1-P(T'|R')=1-0.78=0.22$$

从而

$$P(R \mid T)=\frac{(0.65)(0.73)}{(0.65)(0.73)+(0.35)(0.22)}=\frac{0.4745}{0.4745+0.0770}=\frac{0.4745}{0.5515}=0.86$$

因此，如果随机抽取的销售人员有良好销售记录的概率是 65%，那么在已知选出的人通过测试的条件下，他曾有良好销售记录的概率是 86%.

(b)问题：$P(R' \mid T')$是多少?

利用(9.15)式，

$$P(R' \mid T')=\frac{P(R')P(T' \mid R')}{P(R')P(T' \mid R')+P(R)P(T' \mid R)}$$

唯一剩下的是

$$P(T' \mid R)=1-P(T \mid R)=1-0.73=0.27$$

因此

$$P(R' \mid T')=\frac{(0.35)(0.78)}{(0.35)(0.78)+(0.65)(0.27)}=\frac{0.2730}{0.2730+0.1755}=\frac{0.2730}{0.4485}=0.61$$

因此，如果随机抽取的销售人员有不好销售记录的概率是 35%，那么在已知选出的人未通过测试的条件下，他有不好销售记录的概率是 61%.

(c)问题：$P[(R\cap T)\cup(R'\cap T')]$是多少?

因为 $R\cap T$ 和 $R'\cap T'$ 是互不相容的，由 8.6 节性质 4 知道

$$P[(R\cap T)\cup(R'\cap T')]=P(R\cap T)+P(R'\cap T')$$

而且根据(9.4)式

$$P(R\cap T)=P(R)P(T \mid R)=(0.65)(0.73)=0.4745$$
$$P(R'\cap T')=P(R')P(T' \mid R')=(0.35)(0.78)=0.2730$$

因此

$$P(\text{测试与记录一致})=P[(R\cap T)\cup(R'\cap T')]=0.4745+0.2730=0.75$$

因此，测试结果和原有销售记录一致的概率是 75%.

树型图

9.24　一只罐中有 4 个不同颜色的球，一个红色(R)，一个白色(W)，一个黄色(Y)，一个蓝色(B). 现从罐中任取一个，再取第二个. 利用树型图求至少有一个是黄色球的概率是多少?

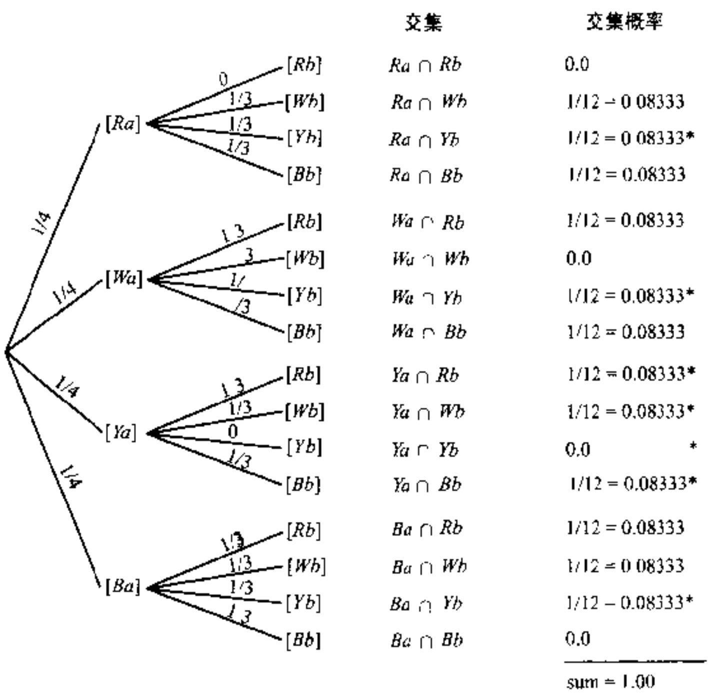


图 9-8

解 ☞ 用 Ra, Wa, Ya 和 Ba 表示第一次抽取的不同结果，用 Rb, Wb, Yb 和 Bb 表示第二次抽取的不同结果，那么这个试验的树型图如图 9-8 所示.

在图 9-8 中带 * 号的是与 P(至少有一个黄球)有关的概率，由于这些事件是互不相容的，由 8.6 节性质 4

$$P(\text{至少有一个黄球}) = P(Ra \cap Yb) + P(Wa \cap Yb) + P(Ya \cap Rb) + P(Ya \cap Wb) + P(Ya \cap Yb) + P(Ya \cap Bb) + P(Ba \cap Yb)$$

因为 $P(Ya \cap Yb) = 0.0$

$$P(\text{至少有一个黄球}) = 6(0.08333) = 0.50$$

9.25 一个高中生想进医学院，但她必须先去一所本地学院或有名的大学. 而这所学院已答应接收她，可她想去大学，但那所大学接收她的概率只有 65%，她又想，若去学院，她有 95%的概率可以毕业，然后有 50%的把握去一所医学院，相反若去大学，她有 70%的概率可以毕业，然后有 75%的把握去医学院，而且她只有从学院或大学毕业后才能进医学院. 利用树型图求 P(被医学院录取).

解 ☞ 用 U 和 C 表示她去大学或学院，用 G 和 G' 表示她是否能毕业，用 M 和 M' 表示她是否被医学院录取，用树型图 9-9 表示如下.

图 9-9 中带 * 号表示的是与 $P(M)$ 有关的概率，因为它们是互不相容的，由 8.6 节性质 4 知，

$$P(M) = P(U \cap G \cap M) + P(U \cap G' \cap M) + P(C \cap G \cap M) + P(C \cap G' \cap M)$$

而且，因为 $P(U \cap G' \cap M) = P(C \cap G' \cap M) = 0$

$$P(M) = P(U \cap G \cap M) + P(C \cap G \cap M) = 0.34125 + 0.16625 = 0.51$$

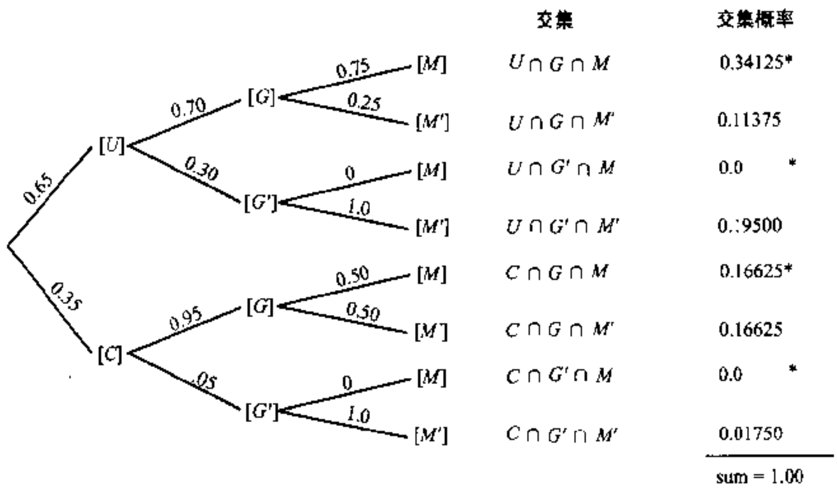


图 9-9

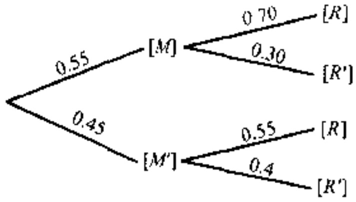
9.26 一位参与议员改选的竞选议员想为持续一分钟的系列电视广告筹资 10 万美元. 他认为有 55%的可能筹到这些钱，假设他筹到了这些钱，就有 70%的机会当选. 如果筹不到这些钱，仍有 55%的可能当选，用树型图求出条件概率 P(已知他当选的条件下，他没有筹到资金).

解 ☞ 如果我们用 M 和 M' 分别表示是否筹集到资金，用 R 和 R' 表示能否重新当选，用树型图 9-10表示如下.

利用(9.15)式和图 9-10 中事件交集概率栏中带 * 号的值来求 $P(M'|R)$，

$$P(M' \mid R) = \frac{P(M')P(R \mid M')}{P(M')P(R \mid M') + P(M)P(R \mid M)} = \frac{P(M' \cap R)}{P(M' \cap R) - P(M \cap R)}$$

$$= \frac{0.2475}{0.2475 + 0.3850} = \frac{0.2475}{0.6325} = 0.39$$



交集	交集概率
$M\cap R$	0.3850*
$M\cap R'$	0.1650
$M'\cap R$	0.2475*
$M'\cap R'$	0.2025
	sum = 1.00

图 9－10

计数法则

9.27　利用**计数法则：乘法原则**来计算下列试验中样本空间的样本点数：(a)连续 11 次投币，(b)从一副有 52 张的纸牌中有放回地连续抽取 5 次，

解 ☞

(a) 这个试验包括 $k=11$ 次连续的试验，每次试验有两种可能的结果，利用(9.16)式，

$$\#\ 样本点数=n_1\times n_2\times\cdots\times n_k=2^{11}=2{,}048$$

(b) 这个试验有 $k=5$ 次连续试验．每次试验有 52 种可能结果，因此

$$\#\ 样本点数=n_1\times n_2\times\cdots\times n_5=52^5=380{,}204{,}032$$

9.28　对于习题 9.27 中的试验，利用(9.9)式和 8.6 节性质 7，可以确定以下事件的概率：(a) P(在 11 次试验中每一次都出现正面)，(b)P(在 5 次试验中每次都抽到红桃 Q)．

解 ☞

(a)如果我们用 $H_1,H_2,\cdots,H_{11}$ 表示在 11 次试验中掷每次后得正面的独立事件，由(9.9)式，

$$P(H_1\cap H_2\cap\cdots\cap H_{11})=P(H_1)P(H_2)\cdots P(H_{11})$$

从而，因为 $P(H_1)=P(H_2)=\cdots=P(H_{11})=\frac{1}{2}$

$$P(H_1\cap H_2\cdots\cap H_{11})=\left(\frac{1}{2}\right)^{11}=0.00049$$

由性质 7 看出，如果 $A=\{$出现 11 次正面$\}$，那么只有一种方式出现这一结果，$N_A=1$．由习题 9.27(a)我们知道 $N=2^{11}=2{,}048$．因此

$$P(A)=N_A\left(\frac{1}{N}\right)=1\left(\frac{1}{2{,}048}\right)=0.00049$$

(b)如果用 $Q_1,Q_2,\cdots,Q_5$ 表示在 5 次抽牌每次都抽取到红桃 Q 的独立事件，那么从(9.9)式

$$P(Q_1\cap Q_2\cap\cdots\cap Q_5)=P(Q_1)P(Q_2)\cdots P(Q_5)$$

从而，因为 $P(Q_1)=P(Q_2)=\cdots=P(Q_5)=\frac{1}{52}$

$$P(Q_1\cap Q_2\cdots\cap Q_5)=\left(\frac{1}{52}\right)^5=0.0000000026$$

从性质 7 知，如果 $A=\{$5 次都取到红桃 Q$\}$，那么只有一种方式出现这一结果，从而 $N_A=1$．由习题 9.27(b)我们知道 $N=52^5=380{,}204{,}032$．因此

$$P(A)=N_A\left(\frac{1}{N}\right)=1\left(\frac{1}{380{,}204{,}032}\right)=0.0000000026$$

9.29　汽车制造商对某种型号的汽车设计了几种部件配置选择，(a) 2 个或 4 个门，(b)6 种颜色(红，黄，蓝，绿，白，银白)中任一种，(c) 配置 AM 或 AM-FM 收音机，(d) 自动换挡或人工换挡．这可能设计出多少种不同型号的汽车？如果每一种型号都描绘在一张卡片上，把所有的卡放在一个容器里，然后随机地从中抽取一张，抽到有 2 个门，人工换挡，装有 AM-FM 收音机的红色型号汽车的概率是多少？

解 ☞　试验是挑选一辆汽车．如果我们把 4 种配置的每一种选择看作一次试验．那么，$n_1=\{$门的选择$\}=2$；$n_2=\{$颜色的选择$\}=6$；$n_3=\{$收音机的选择$\}=2$；$n_4=\{$换挡的选择$\}=2$．因此，利用(9.16)式

可得.

$$\# \text{样本点数} = \# \text{汽车所有可能的型号} = n_1 \times n_2 \times n_3 \times n_4$$
$$= 2 \times 6 \times 2 \times 2 = 48$$

利用 8.6 节性质 7,如果 A={有两个门,人工换挡,AM-FM 收音机的红色汽车}. 那么,$N_A=1$,$N=48$,得

$$P(A) = N_A\left(\frac{1}{N}\right) = 1\left(\frac{1}{48}\right) = 0.021$$

9.30 利用(9.17)式和(9.18)式求把八本书在书架上放成一排,有多少种放法?

解☞ 问题:${}_nP_n$ 为多少?

用 n 来代替(9.17)式中的 r,我们得到

$${}_nP_n = n(n-1)\cdots(n-n-2)(n-n+1) = n(n-1)\cdots(2)(1) = n!$$

同样在(9.18)式中用 n 代替 r 得到

$${}_nP_n = \frac{n!}{(n-n)!} = \frac{n!}{(0)!} = \frac{n!}{1} = n!$$

因此,在这个问题中 $n=8$

$${}_8P_8 = 8! = 8\times7\times6\times5\times4\times3\times2\times1 = 40,320$$

9.31 将一副有 52 张的标准纸牌洗好。从中抽取 3 张,按从左到右的顺序排列在桌子上,顺序恰好为同一花色 J,Q,K 的概率是多少?

解☞ 利用 8.6 节中性质 7,如果记事件 A={同一花色且顺序为 J,Q,K},因为有 4 种花色,所以 $N_A=4$. 利用(9.18)式,

$$N = {}_nP_r = {}_{52}P_3 = \frac{n!}{(n-r)!} = \frac{52!}{(52-3)!} = \frac{52!}{49!} = 52\times51\times50 = 132,600$$

因此

$$P(A) = N_A\left(\frac{1}{N}\right) = 4\left(\frac{1}{132,600}\right) = 0.000030$$

9.32 假如你生活在这样一个国家.在那里,汽车牌照上左边有 3 个字母(在一个牌照上没有重复字母),右边有三个数字(也没有重复)。字母从 26 个字母表中随机选择,数字从 0 到 9 的 10 个数中随机选出。假定所有的牌照都有可能得到,你随机地分配到一个牌照,问得到牌照 ABC012 的概率是多少?

解☞ 利用 8.6 节中性质 7,如果记事件 A={ABC012},那么 $N_A=1$. 如果我们把字母的选择作为第一次试验.数字的选择作为第二次试验.然后用(9.16)式和(9.18)式,

$$N = {}_{26}P_3 \times {}_{10}P_3 = \left[\frac{26!}{(26-3)!}\right]\left[\frac{10!}{(10-3)!}\right]$$
$$= \left(\frac{26!}{23!}\right)\left(\frac{10!}{7!}\right) = (26\times25\times24)(10\times9\times8)$$
$$= (15,600)(720) = 11,232,000$$

因此

$$P(A) = N_A\left(\frac{1}{N}\right) = 1\left(\frac{1}{11,232,000}\right) = 0.000000089$$

9.33 在众议院的立法机关中,有 90 个民主党人和 70 个共和党人,用随机方式从民主党中选出一位多数党领导和多数党领导助理,从共和党中选出一位少数党领导和少数党领导助理,这四个领导位置有多少种排列方式?

解☞ 用(9.18)式,民主党的排列方式有

$${}_nP_r = {}_{90}6P_2 = \frac{n!}{(n-r)!} = \frac{90!}{(90-2)!} = 90\times89 = 8,010$$

共和党的排列方式有

$${}_nP_r = {}_{70}6P_2 = \frac{n!}{(n-r)!} = \frac{70!}{(70-2)!} = 70\times69 = 4,830$$

如果我们把两个民主党领导人的选举作为第一次试验。两个共和党领导人的选举作为第二次试验。然后利用(9.16)式,

$$(\text{领导人排列}) = {}_{90}P_2 \times {}_{70}P_2 = (8,010)(4,830) = 38,688,300$$

9.34 如果在习题 9.33 所述的众议院中，有 6 个民主党人士和 4 个共和党人士来自同一个城市。问来自这个城市的议员选中 4 个领导位置的概率为多少？

解 ☞　利用(9.18)式对于来自这个城市的民主党的领导排列有

$${}_nP_r = {}_6P_2 = \frac{n!}{(n-r)!} = \frac{6!}{(6-2)!} = 6 \times 5 = 30$$

对于这个城市的共和党的领导排列有

$${}_nP_r = {}_4P_2 = \frac{4!}{(4-2)!} = 4 \times 3 = 12$$

然后，像在习题 9.33 计算这 4 个领导的可能配置方式一样，只是现在限制在这个城市的议员之中，

$$(\text{领导人排列}) = {}_6P_2 \times {}_4P_2 = (30)(12) = 360$$

因此，利用 8.6 节性质 7，如果 $A=\{4$ 个领导人来自同一个城市$\}$，$N_A=360$，又根据习题 9.33，$N=38,688,300$，那么

$$P(A) = N_A\left(\frac{1}{N}\right) = 360\left(\frac{1}{38,688,300}\right) = 0.0000093$$

9.35 从单词 CLOVER 无重复地任意抽出三个字母组成一个单词，每个单词写在一张卡片上，把所有的卡放在一个碗中，从中随机抽取一张，抽到以元音字母为首的单词的概率是多大？

解 ☞　利用(9.18)式，从 CLOVER 中任意选取 3 个字母组成的单词总数是

$${}_nP_r = {}_6P_3 = \frac{n!}{(n-r)!} = \frac{6!}{(6-3)!} = 6 \times 5 \times 4 = 120$$

由于在 CLOVER 中有两个元音字母(E,O)，利用(9.16)式可以得到以元音字母为首的三个字母的单词数

$$2 \times {}_4P_2 = 2\left[\frac{4!}{(4-2)!}\right] = 2(4 \times 3) = 24$$

因此，利用 8.6 节中性质 7，若事件 $A=\{$这个学生准备的 4 道题$\}$，则 $N_A=24$，$N=120$，那么

$$P(A) = N_A\left(\frac{1}{N}\right) = 24\left(\frac{1}{120}\right) = 0.20$$

9.36 一个销售商向一顾客展示了 40 套浴具，用(9.20)式，求这位顾客从他商店销售的浴具中购买 5 套共有多少种不同的方式？

解 ☞　问题：${}_{40}C_5$ 是多少？用(9.20)式，

$${}_{40}C_5 = \frac{n!}{r!(n-r)!} = \frac{40!}{5!(40-5)!} = \frac{40 \times 39 \times 38 \times 37 \times 36 \times 35!}{(5 \times 4 \times 3 \times 2 \times 1)(35!)} = 658,008$$

9.37 一个大学历史课的讲师列出 20 道可能问答题，他将从中任选 4 道做为期末试题，由于时间关系，某学生只能准备 4 道题，他所准备的 4 道题就是考题的概率是多少？

解 ☞　利用(9.20)式，从这 20 道题中任意选取 4 道题的总数是

$${}_{20}C_4 = \frac{n!}{r!(n-r)!} = \frac{20!}{4!(20-4)!} = \frac{20 \times 19 \times 18 \times 17 \times 16!}{(4 \times 3 \times 2 \times 1)(16!)} = 4,845$$

因此，用 8.6 节性质 7，若 $A=\{$这个学生准备的 4 道题$\}$，$N_A=1$，$N=4,845$，那么

$$P(A) = N_A\left(\frac{1}{N}\right) = 1\left(\frac{1}{4,845}\right) = 0.00021$$

这学生将以 0.0021 的概率选中这 4 道题。

9.38 在彩票抽奖活动中，你可以从 1 到 42 的数字中任选 6 个数，如果你所选的 6 个数恰好和中奖号码完全一致，你就会获得一等奖，问你获得一等奖的概率？

解 ☞　利用(9.20)式求出从这 42 个数中抽取 6 个数的组合方式

$${}_{42}C_6 = \frac{n!}{r!(n-r)!} = \frac{42!}{6!(42-6)!} = \frac{42 \times 41 \times 40 \times 39 \times 38 \times 37 \times 36!}{(6 \times 5 \times 4 \times 3 \times 2 \times 1)(36!)}$$

$$= 5,245,786$$

因此由 8.6 节性质 7,若 $A=\{$赢得一等奖$\}$,$N_A=1$,$N=5,245,786$,那么

$$P(A)=N_A\left(\frac{1}{N}\right)=1\left(\frac{1}{5,245,786}\right)=0.00000019$$

9.39 在习题 9.38 中,若你所选取的 6 个数有 5 个是中奖号码,可以获得二等奖,问获得二等奖的概率是多大?

解☞ 要想获得二等奖,就必须从 6 个中奖号码中选 5 个,在剩余的 36 个号码中选一个.如果把选 5 个中奖号码作为第一次试验,选一个非中奖号码作为第二次试验,那么利用(9.20)式和(9.16)式,则选对 5 个数的组合总数是

$${}_6C_5\times{}_{36}C_1=\left[\frac{6!}{5!(6-5)!}\right]\left[\frac{36!}{1!(36-1)!}\right]=\left(\frac{6\times5!}{5!1!}\right)\left(\frac{36\times35!}{1!35!}\right)=6\times36=216$$

因此利用 8.6 节性质 7,若 $A=\{$获得二等奖$\}$,$N_A=1$,$N=5,245,786$(见习题 9.38),那么

$$P(A)=N_A\left(\frac{1}{N}\right)=216\left(\frac{1}{5,245,786}\right)=0.000041$$

9.40 如果发给你 5 张牌,其中 2 张方块,2 张红桃,1 张梅花,你有这样一手牌的概率是多大?

解☞ 要得到这样一手牌,必须有 13 张方块中的 2 张,13 张红桃中的 2 张,和 13 张梅花中的一张,如果把上述过程分别看作 3 次试验。利用(9.20)式和(9.16)式,那么得到这样一手牌的所有方式为

$$\begin{aligned}{}_{13}C_2\times{}_{13}C_2\times{}_{13}C_1&=\left[\frac{13!}{2!(13-2)!}\right]\left[\frac{13!}{2!(13-2)!}\right]\left[\frac{13!}{1!(13-1)!}\right]\\&=\left[\frac{13\times12\times11!}{(2\times1)(11!)}\right]\left[\frac{13\times12\times11!}{(2\times1)(11!)}\right]\left[\frac{13\times12!}{1!12!}\right]\\&=\left(\frac{13\times12}{2}\right)\left(\frac{13\times12}{2}\right)\left(\frac{13}{1}\right)=78\times78\times13=79,092\end{aligned}$$

从 52 张牌中选取 5 张牌的所有方式为

$${}_{52}C_5=\frac{52!}{5!(52-5)!}=\frac{52\times51\times50\times49\times48\times47!}{(5\times4\times3\times2\times1)(47!)}=2,598,960$$

因此,由 8.6 节性质 7,若 $A=\{$2 张方块,2 张红桃,1 张梅花的一手牌$\}$,$N_A=79,092$,$N=2,598,960$,那么

$$P(A)=N_A\left(\frac{1}{N}\right)=79,092\left(\frac{1}{2,598,960}\right)=0.030$$

9.41 把 8 本书在书架上摆成一排,这 8 本书能形成多少种不同的组合?

解☞ 问题: ${}_nC_n$ 是多少? 一般利用(9.19)式,

$$\begin{aligned}{}_nC_n&=\frac{n(n-1)\cdots(n-n+2)(n-n+1)}{n!}\\&=\frac{n(n-1)\cdots(2)(1)}{n!}=\frac{n!}{n!}=1\end{aligned}$$

因此,当 $n=8$ 时

$${}_8C_8=1$$

补 充 习 题

条件概率和乘法法则

9.42 从一副新洗好的扑克牌中抽取一张牌.已知抽到的是一张方块的条件下,它是人头牌的概率是多少?

答案 $\frac{3}{13}=0.23$

9.43 从一副新洗好的扑克牌中抽取一张牌,已知抽到的点数是 2,3,4,5 的条件下,它是方块的条件概率是多少?

答案　$\frac{4}{16}=0.25$

9.44　从一副新洗好的扑克牌中抽取一张牌，已知抽取到的是 K 的条件下，它是红色的概率是多少？

答案　$\frac{2}{4}=0.50$

9.45　从一副新洗好的扑克牌中无放回地顺序抽取两张牌，利用(9.4)式，两次都抽到方块的概率是多大？

答案　$\left(\frac{13}{52}\right)\left(\frac{12}{51}\right)=0.059$

9.46　在一个罐中装有 60 个玻璃球弹子，其中 30 个白色的，18 个红色的和 12 个蓝色的，从中无放回地抽取 2 个，两次都抽到蓝色弹子的概率是多大？

答案　$\left(\frac{12}{60}\right)\left(\frac{11}{59}\right)=0.037$

9.47　连续掷两次骰子，两次都没有出现 4 的概率是多大？

答案　$\left(\frac{5}{6}\right)^2=0.69$

9.48　从一副扑克中无放回地连续抽取三张牌，三次都抽到 Q 的概率是多大？

答案　$\left(\frac{4}{52}\right)\left(\frac{3}{51}\right)\left(\frac{2}{50}\right)=0.00018$

9.49　连续掷两次硬币，两次都出现正面的概率是多大？

答案　$\left(\frac{1}{2}\right)^2=0.25$

9.50　从一副扑克中有放回地重复抽取 6 张，6 次均抽到红色的概率是多大？

答案　$\left(\frac{1}{2}\right)^6=0.016$

9.51　去年在某医院的妇产科有 1060 个男孩和 1000 个女孩出生．现在把这些数据作为所有孩子出生的代表，那么在这个妇产科出生的下 4 个孩子都是女孩的概率多大？

答案　$\left(\frac{1,000}{2,060}\right)^4=0.056$

加法法则

9.52　从一副扑克中抽取一张牌，抽到方块或红色的概率是多少？

答案　$\frac{13}{52}+\frac{26}{52}-\frac{13}{52}=\frac{26}{52}=0.50$

9.53　从一副扑克中抽取一张牌，抽到方块或 K 的概率是多少？

答案　$\frac{13}{52}+\frac{4}{52}-\frac{1}{52}=\frac{16}{52}=0.31$

9.54　从一副扑克中抽取一张牌，则抽到 2，3，4 中某一张但不是方块的概率是多少？

答案　$\frac{39}{52}+\frac{12}{52}-\frac{9}{52}=\frac{42}{52}=0.81$

9.55　求 $P(B|A)$当：(a) 若 A，B 互斥，(b)A，B 相互独立．

答案　(a)0，　(b)$P(B)$

9.56　求 $P(A\cap B)$当：(a) 若 A，B 互斥，(b)A，B 相互独立．

答案　(a)0，　(b)$P(A)P(B)$

9.57　求 $P(A\cup B)$当：(a) 若 A，B 互斥，(b)A，B 相互独立．

答案　(a)$P(A)+P(B)$，(b)$P(A)+P(B)-P(A\cap B)=P(A)+P(B)-[P(A)P(B)]$

9.58　100 位美国市长参加一个环境主题会议，其中 50 人是民主党，50 人是共和党，60 人是男性，40 人是女性，且在 60 个男性中有 25 人是民主党．若随机地抽取一名市长，求以下概率：(a)市长是男性民主党人，(b)市长是男性或者是民主党人．

答案　(a)$\frac{25}{100}=0.25$，　(b)$\frac{60}{100}+\frac{50}{100}-\frac{25}{100}=0.85$

9.59　习题 9.58 中的抽取市长试验中，若 40 个女性中有 15 人是共和党，连续两天，每天从这 100 名市长中随机地挑选一名市长(有放回地抽样)，求以下概率：(a)第一天选到男市长，而第二天是选到女市长，(b)第二天，选到男民主党市长或者女共和党市长．

答案　(a)$\frac{60}{100}\times\frac{40}{100}=0.24$，　(b)$\frac{25}{100}+\frac{15}{100}=0.40$

9.60　习题9.58中的抽取市长试验中，25个男性民主党市长中有15人超过45岁，如果从这100名市长中随机选取一名，则选到年龄在45岁以上的男性民主党市长的概率是多少？

答案　$\frac{60}{100}\times\frac{25}{60}\times\frac{15}{25}=0.15$

9.61　见习题9.58，若超过45岁的共有50人(其中30名男市长，20名女市长)，其中25人是民主党，如果从这100人中随机地选取一名，则他是男性或者是民主党的或者在45岁以上的概率是多少？

答案　$\frac{60}{100}+\frac{50}{100}+\frac{50}{100}-\frac{25}{100}-\frac{30}{100}-\frac{25}{100}+\frac{15}{100}=0.95$

9.62　事件A_1,A_2,A_3互斥且是完备的，它们的概率分别为$P(A_1)=0.20$，$P(A_2)=0.60$和$P(A_3)=0.20$.假设$P(B|A_1)=0.10$，$P(B|A_2)=0.50$和$P(B|A_3)=0.40$，求$P(B)$.

答案　0.40

9.63　由问题9.62的事件，求(a)$P(A_1|B)$，(b)$P(A_2|B)$，(c)$P(A_3|B)$.

答案　(a)0.05，(b)0.75，(c)0.20

9.64　事件D_1,D_2和D_3是互不相容且穷举，事件C和C'发生的原因$P(D_1)=0.69$，$P(D_2)=0.05$和$P(D_3)=0.26$，假设$P(C|D_1)=0.10$，$P(C'|D_2)=0.36$，和$P(C'|D_3)=0.54$.那么C'的概率是多少？

答案　0.77

9.65　由习题9.64提供的信息，求C'在D_1，D_2和D_3条件下的概率？换句话说，求(a)$P(D_1|C')$，(b)$P(D_2|C')$，(c)$P(D_3|C')$？

答案　(a) 0.81，(b) 0.04，(c) 0.16

9.66　在复活节，一父亲把一些彩蛋放在篮子中，让他的小女儿找。他把三个篮子分别藏在三个不同的地方，其中，第一个篮子中有两个红色的，第二个篮子中有一个红色的和一个蓝色的。第三个篮子中有两个蓝色的，若他女儿找到一个蓝色的，则她是从第二个篮子中找到的概率是多少？

答案　1/3

9.67　在黑白两个缸中放入一些弹子，黑缸中有12个蓝色弹子和8个红色弹子，从任意一个缸中任取一个弹子，若抽到的弹子是蓝色的，则这是从黑缸中抽出的概率是多少？

答案　2/3

9.68　连续三次掷一枚硬币，利用树型图求至少出现两次反面的概率，用H_1,H_2,H_3分别表示第一，第二，第三次出现正面，T_1,T_2,T_3分别表示第一，第二，第三次出现反面.

答案　树型图如图9-11所示，所有带(*)号是至少出现两次反面的情况，则P(至少出现两次反面)$=0.50$

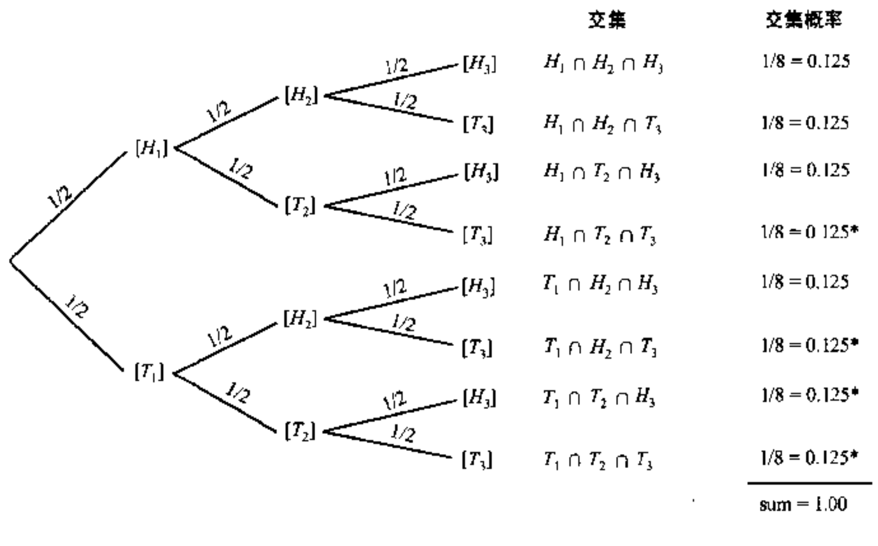
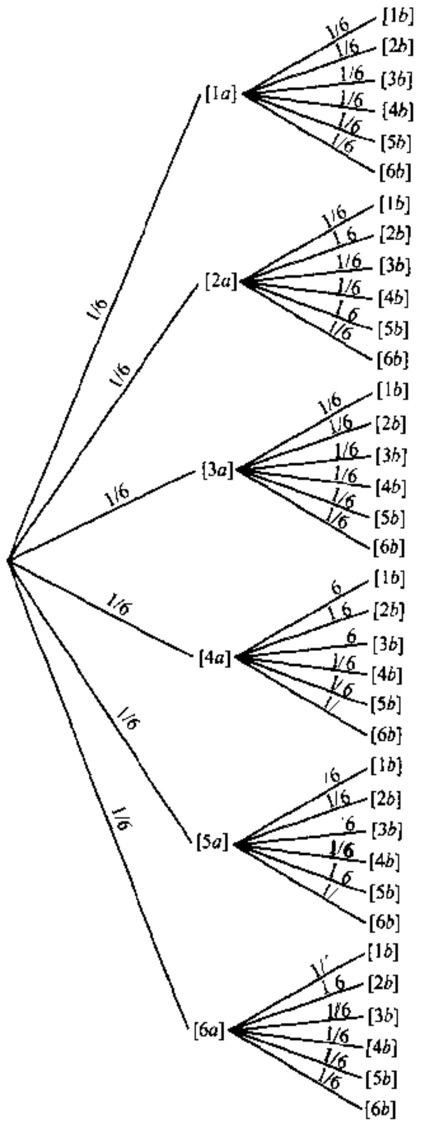


图9-11

9.69　连续两次掷一颗骰子，利用树型图求两次中点数之和是7或11的概率？用$1a,2a,3a,4a,5a,6a$分别表示第一次掷骰子的点数，用$1b,2b,3b,4b,5b,6b$分别表示第二次掷骰子的点数.

答案　树型图如图 9 - 12，所有带（ * ）是点数之和为 7 或 11 的情况。则 P（点数之和是 7 或 11）＝0.22



交集	交集概率
$1a\cap 1b$	1/36 = 0.02778
$1a\cap 2b$	1/36 = 0.02778
$1a\cap 3b$	1/36 = 0.02778
$1a\cap 4b$	1/36 = 0.02778
$1a\cap 5b$	1/36 = 0.02778
$1a\cap 6b$	1/36 = 0.02778*
$2a\cap 1b$	1/36 = 0.02778
$2a\cap 2b$	1/36 = 0.02778
$2a\cap 3b$	1/36 = 0.02778
$2a\cap 4b$	1/36 = 0.02778
$2a\cap 5b$	1/36 = 0.02778*
$2a\cap 6b$	1/36 = 0.02778
$3a\cap 1b$	1/36 = 0.02778
$3a\cap 2b$	1/36 = 0.02778
$3a\cap 3b$	1/36 = 0.02778
$3a\cap 4b$	1/36 = 0.02778*
$3a\cap 5b$	1/36 = 0.02778
$3a\cap 6b$	1/36 = 0.02778
$4a\cap 1b$	1/36 = 0.02778
$4a\cap 2b$	1/36 = 0.02778
$4a\cap 3b$	1/36 = 0.02778*
$4a\cap 4b$	1/36 = 0.02778
$4a\cap 5b$	1/36 = 0.02778
$4a\cap 6b$	1/36 = 0.02778
$5a\cap 1b$	1/36 = 0.02778
$5a\cap 2b$	1/36 = 0.02778*
$5a\cap 3b$	1/36 = 0.02778
$5a\cap 4b$	1/36 = 0.02778
$5a\cap 5b$	1/36 = 0.02778
$5a\cap 6b$	1/36 = 0.02778*
$6a\cap 1b$	1/36 = 0.02778*
$6a\cap 2b$	1/36 = 0.02778
$6a\cap 3b$	1/36 = 0.02778
$6a\cap 4b$	1/36 = 0.02778
$6a\cap 5b$	1/36 = 0.02778*
$6a\cap 6b$	1/36 = 0.02778
	sum = 1.00

图 9 - 12

计数法则

9.70　从 8 个男的和 15 个女的中取一个男的和一个女的组成一个委员会，共有多少种情况？

答案　120

9.71　某学生社团要派遣 4 个学生去参加一个国际会议其中包括一个新生，一个二年级学生，一个三年级的学生，和一个四年级的学生. 现这个社团的志愿者包括 4 个新生，4 个二年级学生，8 个三年级的学生和 3 个四年级的学生。那么从每个年级中选取一个人组成 4 人代表团，共有多少种情况？

答案　384

9.72　一个人准备了 4 条裤子，6 件上衣和 3 条领带去旅行，对这三种服饰，他有多少种不同的穿法？

答案　72

9.73　连续 9 次掷一颗骰子，则样本点是多少？

答案　10,077,696

9.74　见习题 9.73，问在 9 次中均得到 5 点的概率？

答案　0.000000099

9.75　在一次考试中有 10 个选择题，每题有 4 个选项，且只有一个是正确答案，对每个问题都随机地选取一

个答案,则 10 个问题都选对的概率是多少?

答案　0.00000095

9.76　如图 9-4 所示,每条路径代表一个 3 个字母的单词,把每个单词写在一张单独的卡片上,将所有的卡片放在一个碗中,从中随机地选取一张卡片,则它含有字母 W 的概率是多少?

答案　0.75

9.77　一个篮球俱乐部有 18 支球队,联赛结束后,获第一,二,三名的有多少种情况?

答案　4,896

9.78　把 9 名学生分配做 9 门课的助教(一门课一个助教),共有多少种方式?

答案　362,880

9.79　在一小时电视节目中安排 12 条 30 秒的商业信息,有多少种不同方式?

答案　479,001,600

9.80　从 1000 彩票持有者中抽出一等奖和二等奖的获得者,有多少种情况?

答案　999,000

9.81　如果一个销售商要去 5 座城市,那么他有多少条路线?

答案　120

9.82　把 4 节不同的电池放入一个长手电筒,共有多少种方式?

答案　24

9.83　把 200 名申请工作者划分成 4 等:头等、二等、三等、四等,共有多少种情况?

答案　1,552,438,800

9.84　税务机关检查个人收入所得税的情况,若从 100 人中选取 3 人检查,共有多少种情况?

答案　161,700

9.85　一学校要派 5 名学生去参加一个国际会议,但有 30 个人要去,问从中选取 5 人共有多少种方式?

答案　142,506

9.86　一个妇女共有 12 套衣服,现从中选取 4 套,带着去旅行,问她共有多少种选择方式?

答案　495

9.87　若一销售员要去 10 个城市,那么这些城市唯一的组合有几种?

答案　1

第十章 随机变量、概率分布和累积分布函数

10.1 随机变量

这一章主要介绍三个统计中最重要的概念：**随机变量，概率分布和累积分布函数**. 这些函数都有定义域、对应法则和值域(见 1.17 节). 下面先讨论随机变量.

随机变量(也称机会变量)的定义域是样本空间，即随机试验结果的全体(见 8.1 节). 随机变量的对应法则是指定一个且是一个实数值(见 1.17 节)和样本空间中的每个样本点相对应. 它的值域是由对应法则所定义的数值样本空间.

例 10.1 将一枚硬币连掷两次，设随机变量为两次投掷中出现的正面的次数，求(a) 定义域，(b) 对应法则，(c) 值域.

解答

(a)$S=\{$正正，正反，反正，反反$\}$.

(b)对每个样本点统计出现正面的次数.

(c)$S=\{0,1,2\}$.

例 10.2 将一颗骰子连掷两次，设随机变量为两次投掷中出现的点数，求(a)定义域，(b) 对应法则，(c) 值域.

解答

(a)这个试验的样本空间是图 9-12 所示的树型图，有 36 个样本点，

$$S=\{(1,1),(1,2),(1,3),\cdots,(6,4),(6,5),(6,6)\}$$

(b)两次投掷中出现的点数和.

(c)$S=\{2,3,4,5,6,7,8,9,10,11,12\}$.

严格地说，随机变量是以集合论为基础(见 8.6 节)，它是定义在随机试验的样本空间上，取值为实数的函数. 然而，对于随机变量有一个更为常用的意义：如果统计试验的结果是一些定量的测量值(数值)，且这些结果依赖于某种偶然因素，那么所有的测量值变量(见 1.6 节)可以看成是随机变量. 例如，取一群小孩作为随机样本，测量他们的身高，则在常用意义下，这个定量测量使变量可以看作是随机变量.

以上是对随机变量的两种描述，它们的共同点是都能用来处理统计试验得到的数值结果. 常用的意义是用于处理那种从试验得到可能测量值集合. 而集合论方法是用于处理测量过程的本身，函数(或法则)使每一个结果对应一个数. 今后在本书中所讨论的随机变量意义，可以从上下文中看出.

以前讨论的随机变量(见 1.13 节到 1.16 节)用字母表中后面的大写字母 $Z,Y,X,\cdots$ 表示，像以前一样，区别于随机变量，随机变量值用字母表中后面的小写字母 $z,y,x,\cdots$ 表示. 例如 $P(X=x)$ 表示随机变量 X 取值为 x 的概率，如果 $X=1$，那么 $P(X=1)$ 表示随机变量 X 取值为 1 的概率.

例 10.3 指出下列各种情况是否是随机变量：(a)每分钟心脏跳动的次数，(b)掷两次各面上都是 5 的骰子，出现 5 的次数，(c)用秒表示，同一飞机场上，两架飞机起飞的间隔时间，(d)湖水中所含的杀虫剂的量，用每升毫克度量，(e)一个部门每周储备物资的花费，(f)估计物体的大小用三个等级：小，中，大，(g)在 100 个被调查者中对一个问题作肯定回答的人数，(h)在赢得头等奖之前，已经购买的彩票数.

解答

(a)是随机变量.

(b)不是随机变量. 因为试验的结果不是由随机方法所决定；结果总是一对 5，故出现 5 的次数是 2.

(c)是随机变量.

(d)是随机变量.

(e)是随机变量.

(f)不是随机变量.然而,在一个有 50 个物体的样本中,被认为是"小"的物体的个数是随机变量.

(g)是随机变量.

(h)是随机变量.

10.2 离散型与连续型随机变量

随机变量可能是离散的或是连续的,这两个未给的确切定义依赖于随机变量是用常用意义定义还是集合论方法定义.

如果是常用意义定义的随机变量,那么随机变量就是随机确定数的数量测量值变量.因此,在常用意义下,**离散型随机变量**是随机确定的离散数量测量变量,**连续型随机变量**是随机的连续测量变量(见 2.8 节).

如果是用集合论定义随机变量,那么随机变量是定义在随机试验的样本空间上的函数.在这种情况下,**离散型随机变量**的样本空间是**有限的或可数无穷的**,**连续型随机变量**的样本空间是**无限的且是不可数的**.

第 8 章,第 9 章和本章到目前为止所讨论的样本空间都是有限的;它们具有有限多个样本点.例如,样本空间 $S=\{$正正,正反,反正,反反$\}$,它就是有限的,含有 4 个样本点,这样定义在这个样本空间(例如出现正面的次数)上的随机变量是离散型随机变量.类似地连续两次掷一颗骰子,则它的样本空间含有 36 个样本点,设随机变量是两次投掷所得的点数和,它也是离散型随机变量.

在本章中,将第一次考虑可数无穷的样本空间,它有无穷多个样本点,尽管它的样本点是可列的或可数的.例如样本空间是你居住的地方从现在起到发生地震的天数.地震可能在今天,明天,或者在你一生中都没有发生,或者在一百万年以后发生.设随机变量是到发生地震为止的天数,则它是一个能取无限多数的值,从理论上讲,这个天数是可以列出来的,因此是可数的.事实上,这个数列就是所有可能的正整数:$1,2,\cdots,\infty$.因为这个样本空间有可数多个元素,所以这个随机变量是离散型随机变量.

本章将介绍无穷不可数的样本空间,它的元素是不可列或不可数的,也是下册 12 章的主要问题.例如,随机地抽取一群孩子,以英寸为单位,测量他们的身高,这个样本空间包含所有理论上的可能值,假定测量值是完全精确的(见 2.14 节)且测量值是精确到无限位小数.则这个集合是不可列的,因此也是不可数的,因此以实数表示儿童的身高是连续型随机变量.一个无穷且不可数的样本空间对应实数轴上不间断区间中的所有可能值(见 1.20 节).

例 10.4 指出下列各种情况是离散型随机变量,还是连续型随机变量:(a)每分钟心脏跳动的次数,(b)用秒表示在同一机场上,两架飞机起飞的时间间隔,(c)湖水中所含的杀虫剂的量,用每升毫克度量,(d)一个部门每周储备物资的花费,(e)在 100 个被调查者中对问题作肯定回答的人数,(f)在赢得头等奖之前已经购买的彩票数.

解 ☞

(a)离散型随机变量.

(b)连续型随机变量.

(c)连续型随机变量.

(d)离散型随机变量.

(e)离散型随机变量.

(f)离散型随机变量,样本空间是可列的.

10.3 离散型概率分布

我们在 8.6 节定义概率函数为一个数学函数将样本空间中的事件对应于称为概率的实数

且满足概率论的三条公理.这个函数的定义域是样本空间中事件全体,值域是所有这些事件的概率.当样本空间上定义一个离散型随机变量时,概率函数的定义域是随机变量所有可能的取值$(X=x)$,概率函数的值域就是取这些值的概率$[P(X=x)]$.在这种情况下,一个样本空间是由离散型随机变量定义的,概率函数就称为**离散型概率分布或概率散布函数**,这两种说法是相同的.散布函数是指概率散布在随机变量的离散值处.两种都记为$P(X=x)=f(x)$.

离散型概率分布用一个数学公式定义,分布中所有概率都可以通过这个公式计算出来.然而,分布可以用四种形式术语表示:公式本身、列表、图表或由公式计算出来的概率曲线.例如掷一颗骰子试验,所有四种表示如表10.1和图10-1所示,离散型随机变量定义的样本空间是最后骰子向上面的点数,**概率列表**的形式如图10-1(a),这里变量所有值$(X=x)$是已知的,并且给出了取每一个值的概率$[P(X=x)=f(x)]$.同样的信息可以表达成**概率表**,如表10.1的形式.注意表的最后一行$\sum_x f(x)$,它的意思是对随机变量可能取的任意x值求$f(x)$的和.这个符号与$\sum f(x)$有相同的意义,在本书中它们是可以互换的.

表 10.1

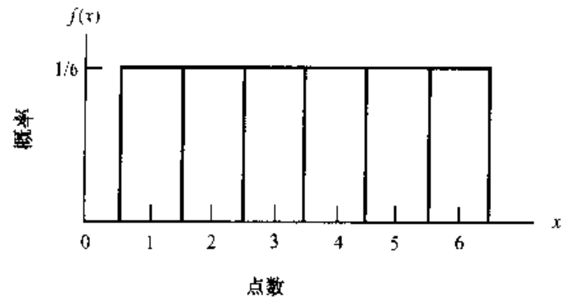
点数 x	概率 $f(x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
	$\sum_x f(x)=1.00$

用曲线图表示离散型随机变量,在本质上和第5章用于表示相对频率分布的曲线是相同的.这样,如果随机变量是离散的比例—水平测量变量,可以用条形图,直方图或棒形图表示概率(见习题5.9和5.10).掷一颗骰子试验的**直方图**如图10-1(b),其中短线表示随机变量的取值$(X=x)$.每一个矩形有等于$f(x)$的高度,底在X轴上,以所取的值为中心.**概率棒形图**如图10-1(c)所示,现在每一个值$(X=x)$用高度为$f(x)$一条垂线表示.在棒形图中可以看到概率散布在x轴上这些离散值上,这就是为什么离散概率分布称为概率散布函数.最后,离散型概率分布用公式表示如图10-1(d).

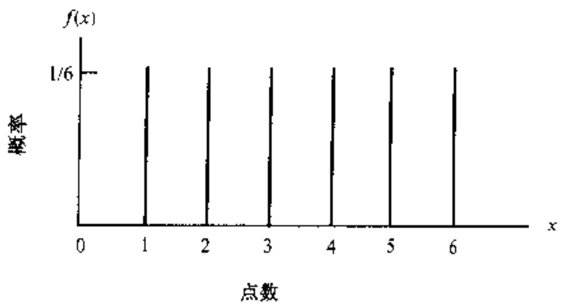
(a) 随机变量的可能值 $x=1, 2, 3, 4, 5, 6$

$P(X=1)=f(1)=1/6$; $P(X=2)=f(2)=1/6$; $P(X=3)=f(3)=1/6$;
$P(X=4)=f(4)=1/6$; $P(X=5)=f(5)=1/6$, $P(X=6)=f(6)=1/6$;

(b)



(c)



(d) $P(X=x)=f(x)=1/6$, $x=1, 2, 3, 4, 5, 6$

图 10-1

由于离散型概率分布是一个概率函数,它服从概率法则和性质(8.6节).这就是说除了个别的以外[$P(X=x)=f(x)$]总是大于等于零[$f(x)\geqslant 0$],对于任何给定离散型概率分布,概率和都等于1[$\sum_x f(x)=P(S)=1.00$;见表10.1].

例10.5　连续三次掷一枚硬币,设随机变量为出现正面的次数,用概率表和概率直方图表示这个离散型概率分布.

解　为了解这个问题,首先,将空间中的简单事件的概率(如图9-11所示的树型图)转化为新的样本空间中的简单事件的概率,其中新的样本空间由随机变量$S=\{0,1,2,3\}$定义.这就意味着我们应求出$f(0),f(1),f(2),f(3)$.

由于所有在树型图上的简单事件(路径)是互不相容的事件,利用集合论性质4(见8.6节),可以得到新的概率值.这样

$$f(0)=P(T_1\cap T_2\cap T_3)=0.125$$

$$f(1)=P(H_1\cap T_2\cap T_3)+P(T_1\cap H_2\cap T_3)+P(T_1\cap T_2\cap H_3)$$

$$=3(0.125)=0.375$$

$$f(2)=P(H_1\cap H_2\cap T_3)+P(H_1\cap T_2\cap H_3)+P(T_1\cap H_2\cap H_3)$$

$$=3(0.125)=0.375$$

$$f(3)=P(H_1\cap H_2\cap H_3)=0.125$$

这个离散型概率分布为如表10.2概率表,它的概率直方图如图10-2.

表10.2

点数 x	概率 $f(x)$
0	0.125
1	0.375
2	0.375
3	0.125
$\sum$	1.00

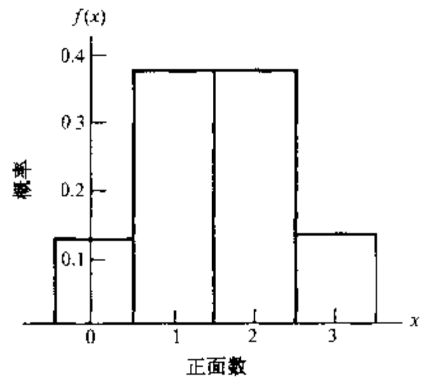


图10-2

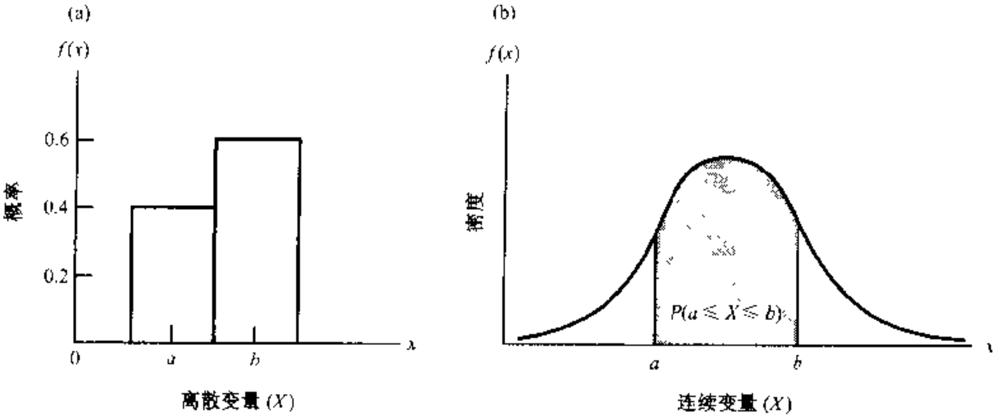
10.4　连续型概率分布

连续型概率分布(或连续型密度函数)对连续型随机变量的样本空间中的事件赋予一个概率,它可以取无限且不可数的特殊值.第12章将主要讨论这种分布,但在这一章我们把它做为概率分布的一般介绍中的一部分.

首先比较连续型概率分布和离散型概率分布的异同点,离散型概率分布曲线和它的重要性质(见10.1,10.2,10.3节)概括如图10-3(a).分布中每一个离散型概率记为$f(x)$,是基于取特殊值x的离散型随机变量X[性质(1),(2),(3)].函数的定义域是X定义的样本空间中的样本点全体($X=x$),值域是这些样本点的概率值[$P(X=x)=f(x)$][性质(4)].离散型概率分布可以表示为列表,图或公式[性质(5)].直方图[如图10-3(a)]中每一个矩形的高度表示概率[$P(X=x)=f(x)$],因此由集合论的公理Ⅰ(见8.6节)可以知道,每一个长方形都不会延伸到X轴以下;即$P(X=x)=f(x)\geqslant 0$[性质(6)].最后,对于离散型概率分布$\sum_x f(x)=P(S)=1.00$总成立[性质(7)];也就是说,任何试验的S中的一个事件,必然发生(见8.6节 公理Ⅱ).而且,对于用直方图表示一个离散概率分布,如果在X轴上相继x值之间的距离总是1.0(如$a=2,b=3$),那么在x值上方每一个长方形的面积是$f(x)$,直方图的总面积是$\sum f(x)=1.0$.

连续型概率分布曲线和它的主要性质概括如图10-3(b).对于连续型概率分布,也记为$f(x)$,基于能取无限不可数的特殊值x的连续型随机变量X[性质(1),(2),(3)].连续型随机

变量 X 定义了有无限不可数多个样本点的样本空间，因此，随机变量取任何一个特殊值 x 的概率都是 0，$[P(X=x)=0$；性质(4)$]$. 由于连续型概率分布不能对每一个样本点赋于一个概率值，它赋予每个样本点一个称为**概率密度**的实数. 这样分布是一个函数其定义域是由 X 定义的样本空间的所有的样本点，值域是所有样本点的 $f(x)$ 值. 这些 $f(x)$ 值对每一个 x 表示概率的密度或浓度的理论测量值. 由于样本点是无限不可数的，相应的概率密度也如此，这样连续型概率分布不能表示为列表或表格的形式，只能表示为连续型曲线[如图 10－3(b)]或者是表示成这条曲线的数学公式[性质(5)].



性质

(1)用 X 表示离散型随机变量

(2)用 x 表示随机变量的特殊值

(3)离散型概率分布(也称作概率分布函数)记作 $f(x)$

(4)$P(X=x)=f(x)$

(5)可以用列表、图表、图、公式表示

(6)$P(X=x)=f(x)\geqslant 0$

(7)$\sum_x f(x)=P(S)=1.00$

性质

(1)连续型随机变量记做 X

(2)随机变量的特殊值记做 x

(3)连续型概率分布(概率密度函数)记做 $f(x)$

(4)$P(X=x)=0$

(5)可以用公式或图表示

(6) $P(a\leqslant X\leqslant b)=\int_a^b f(x)\mathrm{d}x\geqslant 0$ where $f(x\geqslant 0)$

(7) $P(-\infty<X<\infty)=\int_{-\infty}^{\infty} f(x)\mathrm{d}x=P(S)=1.00$

图 10－3

如果 X 轴上方曲线的高度(如图 10－3(b))表示概率密度，而不是概率$[P(X=x)]$，那么连续型概率分布曲线上的概率在哪儿呢？**曲线下的面积**，当利用连续型概率分布计算概率时，这个概率就是 X 取区间$[a,b]$上某些值的概率$[P(a\leqslant X\leqslant b)]$，这个概率就是从 a 到 b 区间上的面积，即是由 X 轴，$x=a$，$x=b$，和曲线 $f(x)$ 围成部分的面积[如图 10－3(b)阴影部分].

为了计算 X 区间$[a,b]$上取值的概率，需要用到**积分运算**. 本书中不需要积分运算，但是如果你熟悉计算，就可以发现性质(6)中 $P(a\leqslant X\leqslant b)=\int_a^b f(x)dx$ 表示为了求 X 在区间$[a,b]$上取值的概率是必须为 $f(x)$ 从 a 到 b 的积分. 关于这个方法的原始想法是，图 10－3(b)中区间$[a,b]$上的垂直的密度线，由于代表基点的一个 x 值它是无限小的，因此这些直线只有高度而没有宽度，所以这些线没有面积. 但是在区间$[a,b]$上无数条这样的线进行叠加(积分)，这个求和过程就产生了一个面积，这个面积就是概率$[P(a\leqslant X\leqslant b)]$. 性质(6)也表明由集合论的公理 I (见 8.6 节)知概率$[P(a\leqslant X\leqslant b)\geqslant 0]$成立，同时 $f(x)\geqslant 0$ 也成立.

由于图 10－3(b)中 a，b 上的垂直的密度线没有面积，因此它们是否包括在从 a 到 b 的区间内，都不影响区间上的面积. 更形式地讲在计算中是否包括区间端点值，并不影响 X 在此区间上取值的概率. 因此，对于连续型随机变量

$$P(a<X\leqslant b)=P(a\leqslant X\leqslant b)=P(a<X<b)=P(a\leqslant X<b) \qquad (10.1)$$

最后，性质(7)表明，从 $-\infty$ 到 $+\infty$ 区间上连续型概率分布以下的总面积(总概率)总是等于 $P(S)=1.00$.　[这里假定在区间$(-\infty,+\infty)$上不在 S 中的 x，$f(x)=0$.]

注意在图 10－3(b)中,Y 轴上没有刻度,可以直接标出 $f(x)$ 值,这是常用的,因为只是概率密度函数以下的面积对于统计方法有重要作用.

对于连续型分布的概率运算,积分运算是必须的,但是对一般统计学者,不要求掌握这门知识的,因为对一些重要概率分布的相关运算都有专门的表格,在本书的附录中提供了一些这样的表格.

10.5　离散型概率分布和描述性分布的关系

由常用定义(见 10.1 节)的随机变量是随机统计试验的所有可能的数字结果的集合;是试验中定量测量变量.因此进行一个这样的试验得到的数据代表随机变量的观察值.这些数据可能是整个测量总体或者从这个总体中抽取的一个样本(见 3.1 和 3.3 节).无论哪种方法,从前几章中知道如何将这些数据能组织成描述性分布(第 4 章),表示为曲线(第 5 章),中心趋势(第 6 章)和离差(第 7 章)的统计度量来描述.

从前面关于总体和样本有关概念(见 3.1 节到 3.5 节)的介绍中可以看出,由于总体不便应用的,总是抽取样本并对它们的总体作统计推断.这些真实数据的样本可以用来**估计**总体的特性,由于总体太大或太分散,这些特征难于测量,或者只是假设。样本信息可利用统计理论,找出一个总体的分布恰当的理论的数学描述或**模型**.事实上概率分布就是总体分布的理论模型;特别是总体相对频数的分布的理论模型.因此离散型概率分布就是离散型随机变量总体相对频率分布的理论模型.

例 10.6　现通过基因工程培养一种新的白鼠,在实验室里有 500 只这样白鼠的自然总体.在它们的白毛上分别点上一个,两个或三个黑点,然后测量这 500 只白鼠身上的点数(离散随机变量).测量总体的频数和相对频数如图 10－4(a)所示.如果从总体中随机地抽取一只白鼠,这只白鼠身上有一个点,两点和三个点的概率分别是多少?

解 ☞　由随机变量所决定的样本空间是 $S=\{1,2,3\}$,有关的概率[由图 10－4(a)中的频数利用古典概率函数(见 8.6 节)计算]是:$P(1)=100/500=0.2$,$P(2)=300/500=0.6$,$P(3)=100/500=0.2$.离散型概率分布的棒型图如图 10－4(b)所示.如果比较图 10－4(a)和 10－4(b)就会发现总体的相对频数分布和离散型概率分布是一样的.对于**真实有限的**测量总体,总体中测量的相对频数也是概率,同时这种测量必须是从总体中随机地抽取.随机地抽取一只白鼠,它身上有一点的概率是 0.2,两个点的概率是 0.6,三个点的概率是 0.2.

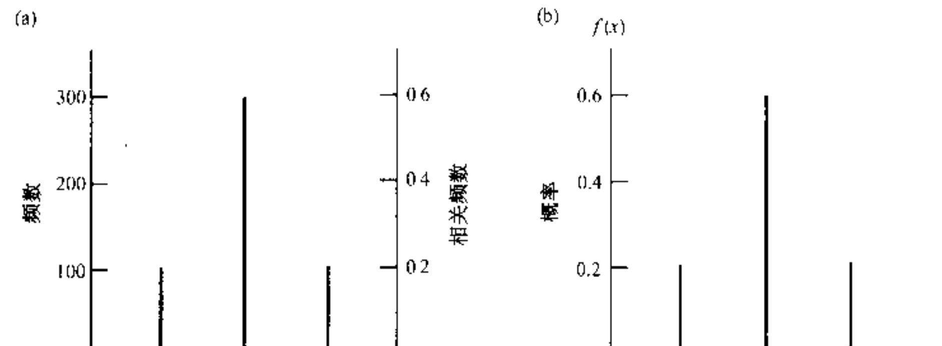


图 10－4

测量总体是真实的如例 10.6 所描述的,有限的且相对较小的总体,然而在大多数的统计分析中不是这种情况.比较典型的是,测量总体不能全部测量或者只是假设的总体.一个假设测量总体的例子就是掷一颗骰子,数它出现点数的各种可能,这个总体不能用于分析,有用的是:(1)来自这个总体的**经验样本**(观测每一次投掷结果),(2)它的相对频数的数学模型:点数这个离散型变量的离散概率分布(如图 10－1).

在统计分析中，如果最合适的概率分布已被选做总体相对频数分布的模型，那么从这个总体中选取的样本相对频数分布，随着样本容量的增大越来越接近概率分布. 从如图8－1掷骰子试验已经看到一点，样本的相对频数估计随着样本容量的增大，越来越接近于理论概率. 对于如图 8－1 所示的掷 240 次掷骰子，图 10－5 给出了六个点数的样本相对频数和理论的离散概率分布的比较.

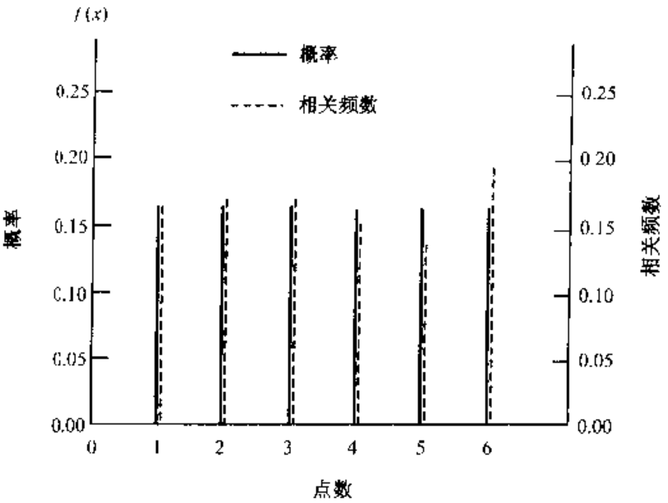


图 10－5

10.6　连续型概率分布和描述性分布的关系

10.5 节指出离散型概率分布是离散型随机变量总体相对频数分布的理论的，数学模型. 类似地，连续型概率分布是连续型随机变量总体相对频数分布的理论的数学模型.

为了理解连续型概率分布的新解释，回到例 10.6 基因工程小白鼠的讨论中. 几年以后，有数以千计的这种白鼠. 将它们看作是现在和将来所有这种老鼠的无限大的假设总体的一个样本. 从这些白鼠中随机地抽取 100 只成年雄鼠，测量它们的体重以克为单位(连续型随机变量). 由于重量是连续性变量，它只是量到最接近的克数. 所得的数据显示为组相对的频数分布(见 1.5 节)，如图 10－6(a)所示. 把数据分成 9 组，间隔是 3 克. 这样结果的直方图和相应的折线(见 5.4 和 5.6 节)是对称的且是单峰的.

将这 100 只白鼠放回总体，重新随机地抽取 500 只成年雄鼠，测量它们的体重并精确到 0.1 克. 把这个结果分成 28 个等重的组(间隔是 1 克)，它的相对频数直方图和相应的折线如图 10－6(b). 曲线同样是对称单峰的.

如果这样抽样过程无限继续下去，且每一次样本容量 n 都增大，测量位取到小数点后更多的位数，且组的宽度越来越细，随着样本容量接近总体容量 N，相对频数折线就会变为一条光滑曲线，称为相对频数曲线，如图 10－6(c)所示. 当 $n=N$ 时，极限达到这个过程，并且在这一点上样本分布变为总体分布. 然而这样的总体分布不便利用，统计理论为它提供连续型概率分布形式的理论模型，这就是为什么连续型概率分布也称作是相对频数分布的**极限形式**；当 $n\to N$，测量位数小数点后更多的位数且宽度加细时，样本相对频数折线就接近于它们的理论曲线.

给总体相对频数分布尽可能选择一个最接近的概率分布模型. 这种选择是概括得到数据的统计试验信息和样本特征的分析. 如果选择的概率分布有比较好的近似，那么就能用它估计总体相对频数. 这是可能的，由于概率分布是随机变量 X 在区间$[a,b]$上取某个值的概率，通过计算它的概率也是对在区间$[a,b]$上总体相对频数的估计. 所以，这种概率也给出 X 的未来样本在区间$[a,b]$上的**期望相对频数**.

如图 10－6(a)和 10－6(b)中的相对频数的折线基本上是对称的钟型曲线，它的尾部趋于

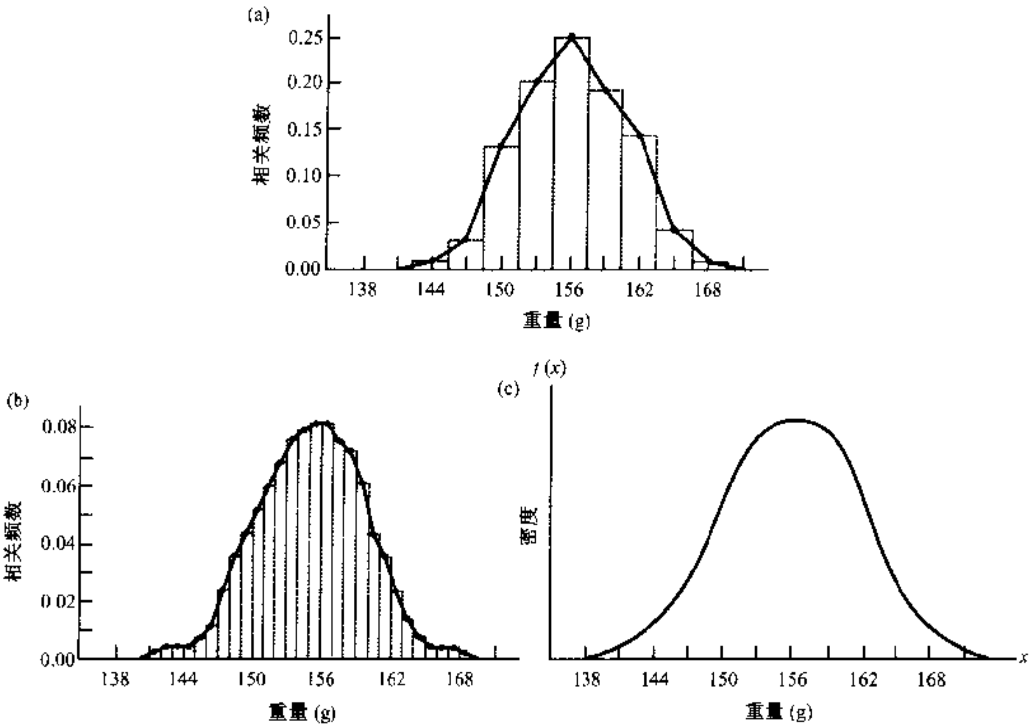


图 10 - 6

X 轴. 这些特征刻划了最重要最常用的连续型概率分布:**正态分布(或正态概率分布,正态概率密度函数)**,从图 10 - 6(c)所示的正态概率分布所示,可以看出它是图 10 - 6(a)和(b)的描述(经验)性分布的很好的近似,由于实际生活中的许多连续型变量的相对频数分布都能用正态分布拟合,所以这个分布将是第 12 章的主要内容,且是本书后部分的一个重点.

10.7　离散型随机变量的累积分布函数

在 10.5 节指出,离散型概率分布(概率散布函数)是离散型随机变量的总体相对频数分布的理论数学模型. 类似地**离散型随机变量累积分布函数(也称分布函数或者累积概率分布)**是变量的总体或者"小于式"累积分布(见 4.9 节和问题 5.26)的理论数学模型. 所以随机变量 X 的总体或者"小于式"累积相对频数分布给出等于或者小于 $X=x$ 值的总体相对频数,累积分布函数给出随机变量 X 取值

$$F(x) = P(X \leqslant x) \tag{10.2}$$

如果 X 是离散型随机变量,对任意实数 a 要求 $P(X \leqslant a)$,那么这个概率可以由公式

$$F(a) = \sum_{x \leqslant a} f(x) \tag{10.3}$$

计算,其中 $\sum_{x \leqslant a} f(x)$ 表示:对所有小于或者等于 a 的 x 值,求 $f(x)$(离散型概率分布)的和.

例 10.7　在掷骰子试验中,随机变量为出现的点数,把表 10 - 1 所示的离散型概率分布变换为累积分布函数,用标准记号概括这个函数,然后作图.

解 ☞　这个试验的累积分布函数可以概括为图 10 - 7(a)或 10 - 7(b),其中图 10 - 7(b)的简化形式是标准形式.

这个概括利用上面的两个公式:对所有的 $x \in (-\infty, +\infty)$, $F(x) = P(X \leqslant x)$. 和对任意实数 a, $F(a) = \sum_{x \leqslant a} f(x)$. 这两个公式指出随机变量 X 取等于或小于实数 a 的 x 值的概率 $F(x)$ 是对所有 $x \leqslant a$ 的 $f(x)$ 的和. 图 10 - 7(a)中,左边 $f(x)$ 的和是对右边区间中任意 x 求和. 这样,最上面一行,由于随机变量不能取到小于或者等于区间 $(-\infty, 1)$ 上的任何值,所以在这个区间上 $F(x)$ 总是 0. 第二行,随机变量 X 只能取一个值 $x=1$,使得它小于或等于任何值,所以在这个区间上 $F(x)$ 等于 $f(1)$ 由表 10.1 知 $f(1)=1/6$. 然后,第三行,X 可以取两个值 $x=1$ 和 $x=2$,它们小于或

等于区间[2,3)上的任何值，所以对这个区间上的任意 x，$F(x)$等于 $f(1)+f(2)=\frac{2}{6}$. 这个累积过程直到最底下一行表明对区间[6,∞)上的任意 x，$F(x)=f(1)+f(2)+f(3)+f(4)+f(5)+f(6)=\frac{6}{6}$. 图 10-7(b)简单地给出，对应于右边区间所有的 x 值，左边是 $F(x)$的算值.

累积分布函数图如图 10-7(c)所示，非常类似第 5 章的离散型变量的(见图 5.32 和 5.33)"小于式"累积曲线. 然而，在那种图中，随机变量值上面的点表示"小于式"累积频数，相对频数，或百分数，现在在随机变量值上方的点表示 $F(x)$：X 取小于或等于 x 值的概率. 由点向右延伸的直线表明，$F(x)$保持常数，直到下一个随机变量值. 例如在 $x=3$ 上的点表明 $F(3)=3/6$，即 X 取值小于或等于 3 的概率是 3/6，在 $x=3.5$ 上面的线段高度表示 $F(3.5)$也等于 3/6，即 X 取值小于或者等于 3.5 的概率是 3/6.

(a)

$$F(x)=\begin{cases} 0 & -\infty < x < 1 \\ f(1)=1/6 & 1 \leqslant x < 2 \\ f(1)+f(2)=2/6 & 2 \leqslant x < 3 \\ f(1)+f(2)+f(3)=3/6 & 3 \leqslant x < 4 \\ f(1)+f(2)+f(3)+f(4)=4/6 & 4 \leqslant x < 5 \\ f(1)+f(2)+f(3)+f(4)+f(5)=5/6 & 5 \leqslant x < 6 \\ f(1)+f(2)+f(3)+f(4)+f(5)+f(6)=6/6 & 6 \leqslant x < \infty \end{cases}$$

(b)

$$F(x)=\begin{cases} 0 & x < 1 \\ 1/6 & 1 \leqslant x < 2 \\ 2/6 & 2 \leqslant x < 3 \\ 3/6 & 3 \leqslant x < 4 \\ 4/6 & 4 \leqslant x < 5 \\ 5/6 & 5 \leqslant x < 6 \\ 6/6 & 6 \leqslant x \end{cases}$$

(c)

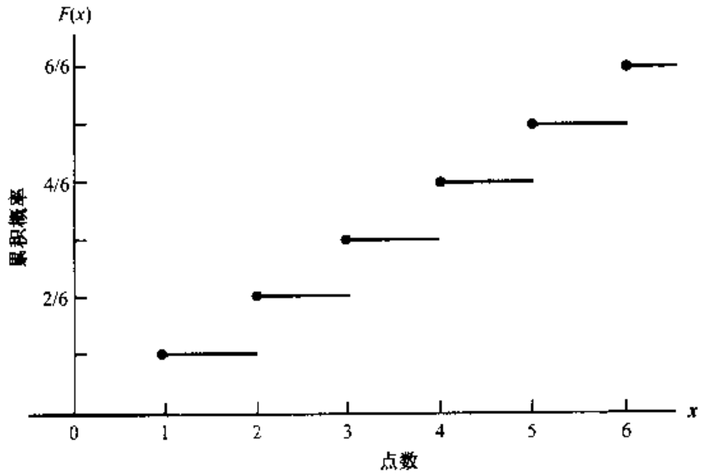


图 10-7

图 10-7(c)表明为什么离散型随机变量的累积分布函数又称为**阶梯函数**：在随机变量每一个取值处 $F(x)$值发生阶梯式变化，每一个随机变量取值($X=x$)处阶梯的大小等于 $P(X=x)=f(x)$，这样在 $x=2$ 和 $x=3$ 之间等于 $P(X=3)=f(3)=1/6$.

由图 10-7(c)还可以看出累积分布另外两条性质：(1)$F(x)$总是大于或者等于 0[$F(x)\geqslant 0$]，(2)随着 $X=x$ 的增大，$F(x)$也增大，当随机变量取最大可能值时 $F(x)=1$.

对于任何离散型变量的累积分布函数 $F(x)$，任意给定的两个实数 a，b 且 $a<b$. 则下列一般规则成立

$$P(a < X \leqslant b) = F(b) - F(a) \tag{10.4}$$

$$P(a \leqslant X \leqslant b) = F(b) - F(a) + f(a) \tag{10.5}$$

$$P(a < X < b) = F(b) - F(a) - f(b) \tag{10.6}$$

$$P(a \leqslant X < b) = F(b) - F(a) + f(a) - f(b) \qquad (10.7)$$

其中 $F(x)$是累积分布函数值,$f(x)$是离散型概率分布值. 习题 10.7,10.8 给出这几个公式的应用.

10.8 连续型随机变量的累积分布函数

连续型随机变量的累积分布函数的定义和离散型随机变量的累积分布函数(见 10.7 节)是相同的. 这样,对于所有实数($-\infty<x<-\infty$),$F(x)=P(X\leqslant x)$. 连续型随机变量的累积分布函数同离散随机变量的累积分布函数的不同之处在于关于和的表示. 对于离散型随机变量,对任意实数 a 通过计算,$F(a)=\sum_{x\leqslant a}f(x)$ 值来求 $P(x\leqslant a)$的值,但对于连续型随机变量,则由计算

$$F(a) = \int_{-\infty}^{a} f(x)dx \qquad (10.8)$$

求 $P(X\leqslant a)$的值.

这个公式用到了 10.4 节介绍的积分计算方法,它表明为计算 $F(a)$只要对 $f(x)$从$-\infty$到 a 积分,图 10-8 说明 $F(a)$即是从$-\infty$到 a 的 X 轴上方的阴影部分,即连续型随机变量 X 取值小于等于 a 的概率. 为了理解这个意思,回忆 10.4 节对连续型分布,只能计算在一个区间上的概率:X 取区间$[a,b]$上某个值的概率,现在是 X 在区间$(-\infty,a]$上取值的概率,为计算它,对区间$(-\infty,a]$上无数条垂直密度线求和(积分).

连续型随机变量的累积分布函数曲线是总体"小于式"相对频数累积曲线的光滑曲线的理论形式,(累积曲线是连续累积分布的图形表示,如 5.11 节描述). 对于 10-8(a)所示的连续型概率分布,光滑的累积曲线如图 10-8(b)所示,其中图 10-8(a)中从∞到 a 区间上密度曲线下面的阴影部分变成X 轴上 a 点处 $F(x)$的高度. 注意由于在连续型概率分布下的总面积(概率)是 1.0,则 $F(x)$向右连续增长到累积概率最大值 1.0.

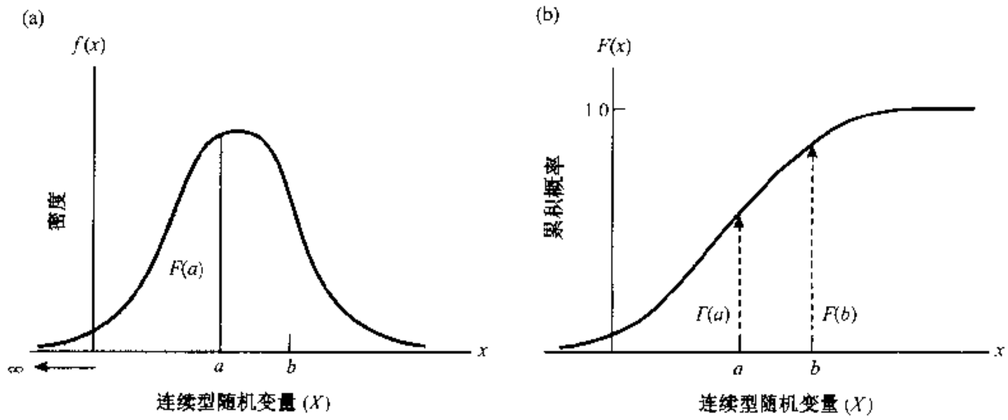


图 10-8

对于连续型随机变量 $P(X=x)=0$ 总成立,因此 $P(a<X\leqslant b)=P(a\leqslant X<b)=P(a<X<b)=P(a\leqslant X\leqslant b)$. 因而,当计算连续型随机变量的累积分布时,关于计算离散型随机变量的累积分布函数 $F(x)$的四个方程[方程(10.4)~(10.7)],都变成$F(b)-F(a)$了. 对于连续型变量的累积分布函数 $F(x)$,及任意两个实数 a,b 且 $a<b$,那么,

$$\begin{aligned} P(a < X \leqslant b) &= P(a \leqslant X \leqslant b) = P(a < X < b) \\ &= P(a \leqslant X < b) = F(b) - F(a) \end{aligned} \qquad (10.9)$$

附录中的概率表就是利用连续型随机变量的这个关系和离散型随机变量的运算法则构造的。

10.9　离散型随机变量的期望值

概率分布有一个**均值**,也称作**期望值(或数学期望,期望)**. 为了理解这个概念,回忆随机变量的概率分布即是它的总体相对频数分布的理论数学模型(见 10.5 节和 10.6 节). 由于相对频数分布可以用统计测度描述,概率分布可以用类似的统计测度描述. 两种类型分布都以均值作为中心趋势度量,两种类型都以方差和标准差作为离差的度量.

离散型随机变量的概率分布的期望值(或均值)定义如下:

设 X 是离散型随机变量,取值为 $x_1, x_2, \cdots, x_k$ 的概率分别为 $f(x_1), \cdots, f(x_k)$,那么 X 的期望值,记为 $E(X)$,且

$$E(X) = \mu = \sum_{i=1}^{k} x_i f(x_i) = \sum_{x} x f(x) \tag{10.10}$$

例 10.8　在例 10.6 挑选小白鼠试验中,离散型随机变量是出现的点数,利用图 10-4(a)中的频数/相对频数和图 10-4(b)所示的离散型概率分布,求 X 的均值.

解　对图 10-4(a)中的总体数据应用不分组的频数分布,求算术平均[(6.11)式],

$$\mu = \frac{\sum_{i=1}^{k} f_i x_i}{N} = \frac{(100\times 1)+(300\times 2)+(100\times 3)}{500} = \frac{100+600+300}{500} = 2.0$$

算术平均公式也可以利用图 10-4(a)的相对频数刻度

$$\mu = \sum_{i=1}^{k}\left[(x_i)\times \frac{f_i}{N}\right] = \sum[(x_i\text{ 的值})\times(\text{总体中 } x_i \text{ 的相对频数})] \tag{10.11}$$

这样,

$$\mu = (1\times 0.2)+(2\times 0.6)+(3\times 0.2) = 2.0$$

对图 10-4(b)中数据,应用求 $E(X)$的公式:

$$E(X) = \sum_{x} x f(x) = (1\times 0.2)+(2\times 0.6)+(3\times 0.2) = 0.2+1.2+0.6 = 2.0$$

有心的读者可能已经发现,在这个例子中,$E(X)$的计算方法和第二种求 μ 的方法是一样的,这是因为像这样一个实际有限的测量总体,这个测量在总体中的相对频数也是从总体中随机抽取这个测量值的概率.

将期望值 $E(X)$作为是均值,因为在多次重复试验中,人们期望这个值就是结果的平均值. 因为 $E(X)$是总体分布的数学模型的均值,经常记为 μ,当需要是区分是随机变量 X 或 Y 的均值时,相应地记为 μ_x 或 μ_y.

离散型随机变量的期望看作是变量的**加权平均**,因为期望值公式就是总体的加权平均公式的一个简化形式. $E(X)$是 X 所有可能取值和以概率为权重的加权平均. 在 6.9 节中总体的加权平均由(6.18)式定义

$$\mu_w = \frac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_i}$$

因此 $E(X)$

$$\mu_w = \mu_x = \mu = \frac{\sum_{i=1}^{k} x_i f(x_i)}{\sum_{i=1}^{k} f(x_i)} \tag{10.12}$$

然而,由于经常遇到的是完全概率分布的期望值,因此上式的分母是所有的概率和,它总是 1,所以公式就变为(10.10)的形式.

同样可以计算,**离散型随机变量函数的期望值**. 如果 X 是离散型随机变量,取值为 $x_1, \cdots, x_k$ 的,概率分别为 $f(x_1), \cdots, f(x_k)$,又 $g(X)$是 X 的函数,那么 $g(X)$的期望值 $E[g(X)]$是

$$E[g(X)] = \sum_{i=1}^{k} g(x_i) f(x_i) = \sum_{x} g(x) f(x) \tag{10.13}$$

总之,随机变量函数的期望值就等于在 x_i 的函数值与 $X=x_i$ 的概率的乘积的和.

例 10.9　求下列离散型随机变量函数的期望值:(a) X^2,　(b) $a+bX$ 其中 a, b 是常数,

(c) $(X-a)^2$, a 是常数.

解

(a) $E(X^2)=\sum_{i=1}^{k}x_i^2 f(x_i)=\sum_x f(x)$

(b) $E(a+bX)=\sum_{i=1}^{k}(a+bx_i)f(x_i)=\sum_x(a+bx)f(x)$

这个式子可以利用和式记号变化公式(见 1.22 节,及习题 1.41 到 1.46)简化为:

$$E(a+bX)=\sum_x(a+bx)f(x)=\sum_x af(x)+\sum_x bxf(x)$$

$$a\sum_x f(x)+b\sum_x xf(x)$$

由于 $\sum_x f(x)=1$ (见 10.3 节)和 $\sum_x xf(x)=E(X)$ [10.10 式]

$$E(a+bX)=a+bE(X)$$

(c) $E[(X-a)^2]=\sum_x(x-a)^2 f(x_i)=\sum_x(x-a)^2 f(x)$

仍利用求和号.

$$E[(X-a)^2]=\sum_x(x-a)^2 f(x)=\sum_x(x^2-2ax+a^2)f(x)$$

$$=\sum_x x^2 f(x)-\sum_x 2axf(x)+\sum_x a^2 f(x)$$

$$=\sum_x x^2 f(x)-2a\sum_x xf(x)+a^2\sum_x f(x)$$

由于 $\sum_x x^2 f(x)E(X^2)$, $\sum_x xf(x)=E(X)$, 和 $\sum_x f(x)=1$,

$$E[(X-a)^2]=E(X^2)-2aE(X)+a^2$$

10.10 连续型随机变量的期望值

连续型随机变量的概率分布的期望值 $E(X)$ 和离散型变量的期望值 $E(X)$ 是相似的,唯一的区别是离散型期望用和式定义,而连续型期望是用积分运算定义.

如 X 是连续型随机变量,密度函数是 $f(x)$,那么 X 的期望的定义是:

$$E(X)=\mu_x=\mu=\int_{-\infty}^{\infty}xf(x)dx \tag{10.14}$$

注意,类似于离散型变量的期望 $E(X)$ 的解释,连续型变量的期望也看作是连续型变量 X 和它连续型概率分布 $f(x)$ 的加权平均($\mu_x=\mu$).

再次强调,本书不要求微积分知识,不要求用本章或别章中牵涉到微积分的公式去计算,而是用书中给出计算结果的公式计算.

10.11 离散型随机变量的方差和标准差

离散性的一个最重要量度就是**离散型随机变量的方差**,离散型随机变量 X 的方差是 **X 与它均值偏差的平方的期望值**,

$$\mathrm{Var}(X)=\sigma_x^2=\sigma^2=E[(X-E(X))^2]=E[(X-\mu)^2] \tag{10.15}$$

由 10.9 节知 $E[g(X)]=\sum_x g(x)f(x)$,故可以定义方差如下:设 X 是离散型随机变量,分别以概率 $f(x_1), f(x_2), \cdots, f(x_k)$ 取值 $x_1, x_2, \cdots, x_k$,那么 X 的方差

$$E[(X-\mu)^2]=\sigma^2=\sum_{i=1}^{k}(x_i-\mu)^2 f(x_i)=\sum_x(x-\mu)^2 f(x) \tag{10.16}$$

在 7.9 节指出测量总体的标准差是这些测量值的方差的平方根(见 7.20 式)即

$$\sigma=\sqrt{\sigma^2}$$

对于**离散型随机变量** X(概率分布)的标准差仍然有

$$\sigma=\sqrt{\sigma^2}=\sqrt{\sum_x(x-\mu)^2 f(x)} \tag{10.17}$$

例 10.10 例 10.6 挑选白鼠试验中,随机变量是点数,求如图 10-4(a)所示的频数/相对频数分布的方差和图 10-4(b)所示离散型概率分布的方差.

解 用不分组频数分布方差公式计算频数/相对频数分布的方差,在 7.2 节中已经得到频数分布的标准差公式(等式 7.29).

$$\sigma=\sqrt{\frac{\sum_{i=1}^{k}f_i(x_i-\mu)^2}{N}}$$

所以

$$\sigma^2=\frac{\sum_{i=1}^{k}f_i(x_i-\mu)^2}{N} \tag{10.18}$$

把图 10-4(a)中的数据代入公式中,又由例 10.8 知 $\mu=\frac{\sum_{i=1}^{k}f_ix_i}{N}=2.0$,得到

$$\sigma^2=\frac{\sum_{i=1}^{k}f_i(x_i-\mu)^2}{N}=\frac{100(1-2.0)^2+300(2-2.0)^2+100(3-2.0)^2}{500}$$
$$=\frac{100+0+100}{500}=0.4$$

这个方差公式也可以写成利用图 10-4(a)中,相对频数刻度形式

$$\sigma^2=\sum_{i=1}^{k}\left[(x_i-\mu)^2\times\left(\frac{f_i}{N}\right)\right]=\sum[(x_i\text{ 减去 }\mu)^2\times(\text{差的相关频数})]$$
$$=[(1-2.0)^2\times(0.2)]+[(2-2.0)^2\times(0.6)]+[(3-2.0)^2\times(0.2)]=0.4 \tag{10.19}$$

把图 10-4(b)中数据代入(10.6)式,再由例 10.8 中 $E(X)=\mu=2.0$ 有

$$\sigma^2=\sum_x(x-\mu)^2f(x)$$
$$=[(1-2.0)^2\times(0.2)]+[(2-2.0)^2\times(0.6)]+[(3-2.0)^2\times(0.2)]$$
$$=0.4$$

就像均值一样,对于方差来说,相对频数的公式和概率公式是一样的,同样因为在一个真实有限的总体中,相对频数和概率相等.

10.12 离散型随机变量的方差和标准差的计算公式

在第 6 章和第 7 章中对于统计度量,定义公式和计算公式是有差别的,这样由 7.12 节知,总体方差的相对频数分布定义式[等式(10.18)],

$$\sigma^2=\frac{\sum_{i=1}^{k}f_i(x_i-\mu)^2}{N}$$

能改写成计算公式[(7.31)式的平方]

$$\sigma^2=\frac{\sum_{i=1}^{k}f_ix_i^2}{N}-\mu^2 \tag{10.20}$$

而计算公式的相对频数形式为

$$\sigma^2=\sum_{i=1}^{k}\left[x_i^2\left(\frac{f_i}{N}\right)\right]-\mu^2 \tag{10.21}$$

对于概率分布有类似的公式.这样由离散型概率分布的方差的定义式可以导出一个计算公式,(10.16)式等于

$$\sigma^2=\sum_x(x-\mu)^2f(x)=\sum_x(x^2-2x\mu+\mu^2)f(x)$$
$$=\sum_x x^2f(x)-\sum_x 2x\mu f(x)+\sum_x\mu^2f(x)$$
$$=\sum_x x^2f(x)-2\mu\sum_x xf(x)+\mu^2\sum_x f(x)$$

注意 $\sum_x f(x)=1$ 和 $\sum_x xf(x)=\mu$

$$\sigma^2=\sum_x x^2f(x)-2\mu^2+\mu^2=\sum_x x^2f(x)-\mu^2$$

注意 $E(X)=\mu$ 和 $E(X^2)=\sum_x x^2 f(x)$，得到计算公式

$$\sigma^2=\sum_x x^2 f(x)-\mu^2=E(X^2)-[E(X)]^2 \tag{10.22}$$

利用例 10.10 所述的计算公式，其中 $E(x)=\mu=2.0$

$$\sigma^2=\sum_x x^2 f(x)-\mu^2=[(1\times 0.2)+(4\times 0.6)+(9\times 0.2)]-(2.0)^2$$

这个结果和例 10.10 利用定义式的结果一样.

因此标准差的计算公式是

$$\sigma=\sqrt{\sum_x x^2 f(x)-\mu^2=E(X^2)-[E(X)]^2} \tag{10.23}$$

例 10.11 在掷骰子试验中，设随机变量是出现的点数(见表 10.1)，分别利用定义式和计算公式求它的方差.

解 首先计算随机变量的期望值

$$\begin{aligned}E(X)&=\sum_x xf(x)=\left(1\times\frac{1}{6}\right)+\left(2\times\frac{1}{6}\right)+\left(3\times\frac{1}{6}\right)\\&\quad+\left(4\times\frac{1}{6}\right)+\left(5\times\frac{1}{6}\right)+\left(6\times\frac{1}{6}\right)\\&=(0.166667)+(0.333333)+(0.5)+(0.666667)+(0.833333)+(1.0)\\&=3.5\end{aligned}$$

那么 $E(X)=\mu=3.5$. 把表 10.1 中的这个变量的概率分布代入定义式[(10.16)式]得

$$\begin{aligned}\sigma^2&=\sum_x (x-\mu)^2 f(x)\\&=\left[(1-3.5)^2\times\frac{1}{6}\right]+\left[(2-3.5)^2\times\frac{1}{6}\right]+\left[(3-3.5)^2\times\frac{1}{6}\right]\\&\quad+\left[(4-3.5)^2\times\frac{1}{6}\right]+\left[(5-3.5)^2\times\frac{1}{6}\right]+\left[(6-3.5)\times\frac{1}{6}\right]\\&=\frac{6.25}{6}+\frac{2.25}{6}+\frac{0.25}{6}+\frac{0.25}{6}+\frac{2.25}{6}+\frac{6.25}{6}=\frac{17.50}{6}=2.92\end{aligned}$$

将子样数据代入计算式(10.22)式，有

$$\begin{aligned}\sigma^2&=\sum_x x^2 f(x)-\mu^2\\&=\left[\left(1\times\frac{1}{6}\right)+\left(4\times\frac{1}{6}\right)+\left(9\times\frac{1}{6}\right)+\left(16\times\frac{1}{6}\right)+\left(25\times\frac{1}{6}\right)+\left(36\times\frac{1}{6}\right)\right]-(3.5)^2\\&=\frac{91}{6}-12.25=2.92\end{aligned}$$

10.13 连续型随机变量的方差和标准差

本章前几节，介绍了离散型随机变量的方差 σ^2 和标准差 σ，现在考虑连续型随机变量的方差和标准差，对离散型变量是用和式定义的.

$$\mathrm{Var}(X)=\sigma_x{}^2=\sigma^2=E[(X-\mu)^2=\sum_x (x-\mu)^2 f(x)$$

和

$$\sigma=\sqrt{\sigma^2}$$

而对连续型变量是用积分式定义：

如果 X 是连续型随机变量，其密度函数为 $f(x)$，那么它的方差就定义为

$$\mathrm{Var}(X)=\sigma_x^2=\sigma^2=E[(X-\mu)^2]=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx \tag{10.24}$$

标准差为

$$\sigma=\sqrt{\sigma^2}$$

10.14　Chebyshev 定理和经验法则

Chebyshev 定理描述的是分布的标准差和取值集中在分布均值之间的关系(见 7.15 节).此定理的另一种形式是：

对任意值 $k \geqslant 1$，随机变量 X 的均值为 μ 标准差是 σ，那么 X 在区间 $\mu \pm k\sigma$ 上取值的概率至少是 $1-\frac{1}{k^2}$.

例 10.12　连续掷骰子两次，设离散型随机变量为两次掷骰子的总点数，其概率分布如表 10.3，(表的由来见习题 10.2)求随机变量在区间 $\mu \pm 2\sigma$ 上取值的概率?

表 10.3

点数 x	概率 $f(x)$	点数 x	概率 $f(x)$	点数 x	概率 $f(x)$
2	0.02778	6	0.13890	10	0.08334
3	0.05556	7	0.16668	11	0.05556
4	0.08334	8	0.13890	12	0.02778
5	0.11112	9	0.11112	$\sum$	1.00

解 ☞　一个解法就是应用 Chebyshev 定理($k=2$)，那么概率至少是

$$1-\frac{1}{k^2}=1-\frac{1}{4}=\frac{3}{4}=0.75$$

对这个问题的精确解法是求 X 在区间 $\mu \pm 2\sigma$ 的所有可能值的概率和. 如表 10.4 所示，此变量的均值和标准差是 $\mu=7.00$ 和 $\sigma=2.41$(见习题 10.15)因此，

$$\mu \pm 2\sigma \text{ 是 } 7.00 \pm 2(2.41), \quad \text{或 } 7.00 \pm 4.82, \quad \text{或 } 2.18 \text{ 到 } 11.82$$

在这个区间上 X 可以取值 3 到 11，由表 10.3 和表 10.4 所示的概率分布，随机变量取这些值中的任何一个的概率是

$$\sum_{x=3}^{11} f(x) = 0.94452, \quad \text{或 } 0.94$$

表 10.4

点数 x	x^2	概率 $f(x)$	$xf(x)$	$x^2f(x)$
2	4	0.02778	0.05556	0.11112
3	9	0.05556	0.16668	0.50004
4	16	0.08334	0.33336	1.33344
5	25	0.11112	0.55560	2.77800
6	36	0.13890	0.83340	5.00040
7	49	0.16668	1.16676	8.16732
8	64	0.13890	1.11120	8.88960
9	81	0.11112	1.00008	9.00072
10	100	0.08334	0.83340	8.33400
11	121	0.05556	0.61116	6.72276
12	144	0.02778	0.33336	4.00032
$\sum$		1.00	7.00056	54.83772

$$\sigma^2=\sum x^2 f(x)-\mu^2=54.83772-(7.00056)^2=54.83772-49.00784=5.82988, \text{或 } 5.83$$

$$\sigma=\sqrt{\sigma^2}=\sqrt{5.82988}=2.41451, \text{或 } 2.41$$

经验方法也是描述标准差和数值在均值附近集中度之间的关系. 对连续型正态概率分布此准则是：

随机变量 X 的均值是 μ 标准差是 σ，它的概率分布近似于正态分布，那么 X 在区间 $\mu \pm \sigma$ 上取值概率约为 0.68，在区间 $\mu \pm 2\sigma$ 上取值概率约为 0.95，在区间 $\mu \pm 3\sigma$ 上取值概率约为1.00.

例 10.13　连续两次掷一颗骰子，设随机变量是出现的点数(见例 10.12)，若把它看成一个连续型的随机变量(见习题 5.9 和 5.26)并假设它的概率分布是近似正态分布(单峰，

钟型且基本对称的),利用经验法则求此随机变量在区间 $\mu \pm 2\sigma$ 上的概率?

解 ☞ 从例 10.12 可以看出,精确的概率为 0.94,由经验准则得到的概率约为 0.95,而由 Chebyshev 定理只能知道概率至少是 0.75.

习 题 解 答

随机变量和概率分布

10.1 指出下列是否为随机变量,若是请说明是离散型的还是连续型的:(a)确定飞机是否准时到达,(b)按种类给鸟分类,(c)申请去医学院的女学生的年龄,(d)加里福尼亚在六个月内测量到地震的次数,(e)取 5 张扑克为一手,在拿到一手 4 张 Q 以前,的手数,(f)风速每小时的英里数,(g)在 20 个问题的测验中,答对的问题个数,(h) 400 个西瓜中每个瓜的重量(以克为单位).

解 ☞

(a) 不是一个随机变量,若考虑 200 架飞机中准时到达的数量,则它就是离散型随机变量.

(b)不是随机变量,但若考虑 80 只鸟中麻雀的数量,则就是离散型随机变量.

(c)是连续型随机变量.

(d)是离散型随机变量.

(e)是离散型随机变量,并可取可数个值.

(f)是连续型随机变量.

(g)是离散型随机变量.

(h)是连续型随机变量.

10.2 连续掷骰子,设随机变量为两次出现的总点数,用概率列表和概率棒形图表示它的概率分布.

解 ☞ 为解决这个问题,首先把样本空间中的简单事件概率(如图 9-12)变换为新的样本空间中的简单事件的概率,其中新的样本空间为 $S=\{2,3,4,5,6,7,8,9,10,11,12\}$ 这意味着必须求出 $f(2)$, $f(3)$, $\cdots$, $f(11)$, $f(12)$

由于树型图 9-12 中所有简单事件是互不相容的,因此可利用集合论中的性质 4(见 8.6 节)求新的概率值,这样

$$f(2)=P(1a\cap 1b)=0.02778$$

$$f(3)=P(1a\cap 2b)+P(2a\cap 1b)=2(0.02778)=0.05556$$

$$f(4)=P(1a\cap 3b)+P(2a\cap 2b)+P(3a\cap 1b)=3(0.02778)=0.08334$$

$$f(5)=P(1a\cap 4b)+P(2a\cap 3b)+P(3a\cap 2b)+P(4a\cap 1b)$$
$$=4(0.02778)=0.11112$$

$$f(6)=P(1a\cap 5b)+P(2a\cap 4b)+P(3a\cap 3b)+P(4a\cap 2b)+P(5a\cap 1b)$$
$$=5(0.02778)=0.13890$$

$$f(7)=P(1a\cap 6b)+P(2a\cap 5b)+P(3a\cap 4b)+P(4a\cap 3b)+P(5a\cap 2b)$$
$$+P(6a\cap 1b)=6(0.02778)=0.16668$$

$$f(8)=P(2a\cap 6b)+P(3a\cap 5b)+P(4a\cap 4b)+P(5a\cap 3b)+P(6a\cap 2b)$$
$$=5(0.02778)=0.13890$$

$$f(9)=P(3a\cap 6b)+P(4a\cap 5b)+P(5a\cap 4b)+P(6a\cap 3b)$$
$$=4(0.02778)=0.11112$$

$$f(10)=P(4a\cap 6b)+P(5a\cap 5b)+P(6a\cap 4b)=3(0.02778)=0.08334$$

$$f(11)=P(5a\cap 6b)+P(6a\cap 5b)=2(0.02778)=0.05556$$

$$f(12)=P(6a\cap 6b)=0.02778$$

这个离散型概率分布的表格形式如表 10.3,棒形图如图 10-9.

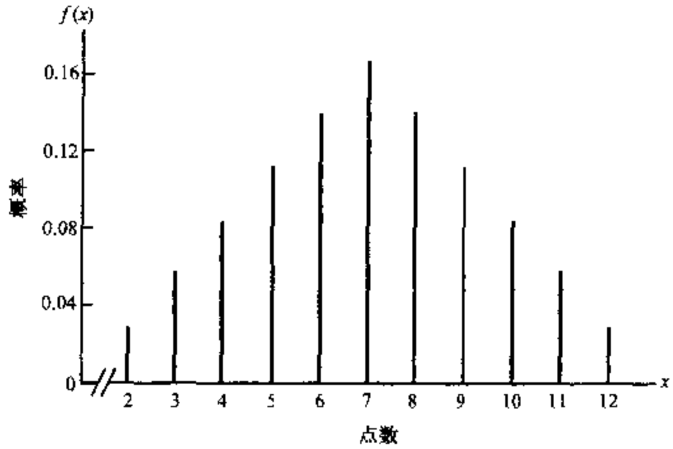


图 10-9

10.3 见例 9.1 和表 9.1,从参与感冒细菌感染研究的 160 个人中无放回的随机选出两个人,随机变量为在一年内被选者中得感冒的人数,求这个离散型概率分布的概率表.

解 用 c_1、c_2 表示第一次,第二次所选人患感冒,用 n_1,n_2 代表第一次,第二次候选人未患感冒,试验的样本空间的树型图如图 10-10(a),(b)所示.

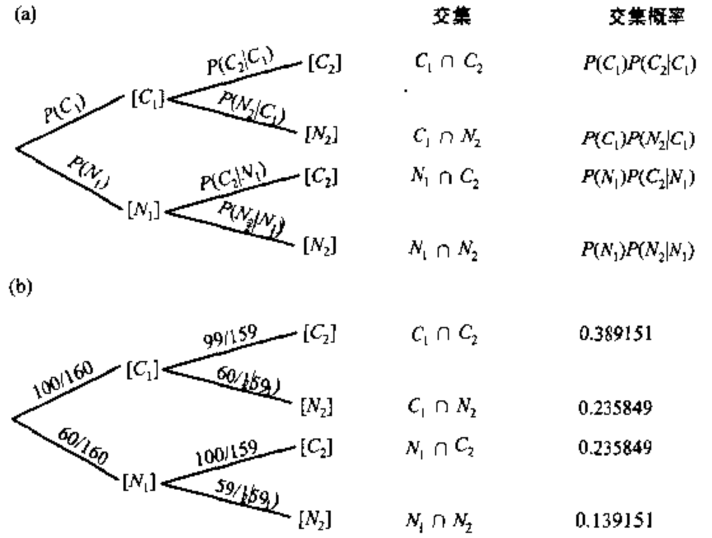


图 10-10

现在把此样本空间中的概率变换为新的样本空间中的概率,这个新的样本空间 $S=\{0,1,2\}$,这意味着必须求出 $f(0)$,$f(1)$和 $f(2)$,仍然应用集合论的性质 4(见 8.6 节).

这样

$$f(0)=P(N_1\cap N_2)=0.139151$$

$$f(1)=P(C_1\cap N_2)+P(N_1\cap C_2)=2(0.235849)=0.471698$$

$$f(2)=P(C_1\cap C_2)=0.389151$$

离散型概率分布如表 10.5 所示.

表 10.5

感冒人数 x	概率 $f(x)$	感冒人数 x	概率 $f(x)$
0	0.139151	2	0.389151
1	0.471698	$\sum$	1.00

10.4 解释下列为什么不是离散型概率分布:(a)$P(X=小)=f(小)=0.5$,$P(X=大)=f(大)=0.5$;(b)$P(X=0)=f(0)=0.2$,$P(X=1)=f(1)=0.6$,$P(X=2)=f(2)=0.3$;(c)$P(X=3)=f(3)=-0.2$,$P(X=4)=f(4)=0.8$.

解 ☞

(a)在 10.1 节中指出随机变量取值必须是实数,不能是"小"或"大",所以它不是随机变量. 又因为离散型概率分布是定义在随机变量的基础之上(见 10.3 节),因此它也不是概率分布.

(b)在 10.3 节中指出每一个离散型概率分布,都有 $\sum_x f(x)=1.00$,但这里 $\sum_x f(x)=1.1$,所以它不是离散型概率分布.

(c)在 10.3 节中指出,对于随机变量 X 的所有可能值 x 都有 $f(x)\geqslant 0$ 恒成立,而这里 $f(3)=-0.2$,因此它不是离散型概率分布.

累积分布函数

10.5 连续掷两次骰子,设随机变量为两次掷骰子中所出现的点数,将表 10.3 中离散型概率分布变换为累积分布函数,用简化形式表示累积分布函数并作图.

解 ☞　累积分布函数和图分别由 10-11(a)和 10-11(b)给出.

(a)

$$F(x)=\begin{cases} 0 & x<2 \\ 0.02778 & 2\leqslant x<3 \\ 0.08334 & 3\leqslant x<4 \\ 0.16668 & 4\leqslant x<5 \\ 0.27780 & 5\leqslant x<6 \\ 0.41670 & 6\leqslant x<7 \\ 0.58338 & 7\leqslant x<8 \\ 0.72228 & 8\leqslant x<9 \\ 0.83340 & 9\leqslant x<10 \\ 0.91674 & 10\leqslant x<11 \\ 0.97230 & 11\leqslant x<12 \\ 1.00008 & 12\leqslant x \end{cases}$$

(b)

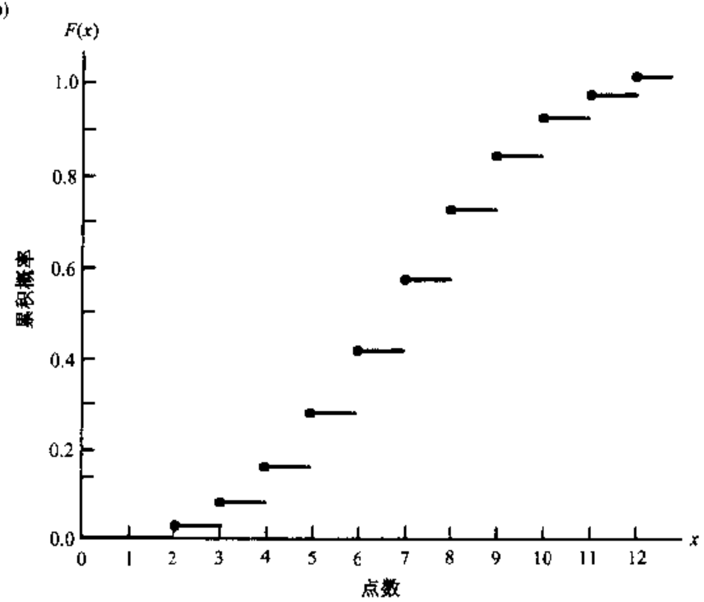


图 10-11

10.6 在感冒病毒研究试验中,设随机变量为感冒人数,将表 10.5 中离散型概率分布变换为累积分布函数,用简化形式表示累积分布函数,且作图.

解 ☞　累积分布函数和图分别由 10-12(a)和 10-12(b)给出.

(a)

$$F(x)=\begin{cases}0 & x<0\\ 0.139151 & 0\leqslant x<1\\ 0.610849 & 1\leqslant x<2\\ 1.000000 & 2\leqslant x\end{cases}$$

(b)

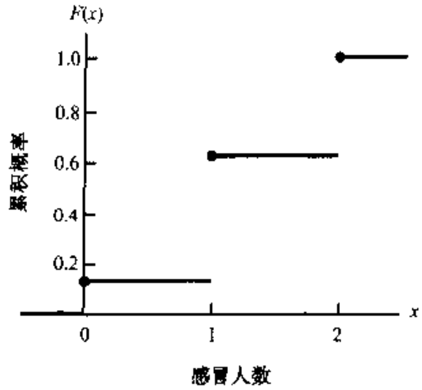


图 10-12

10.7 根据图 10-13 所示的累积分布函数曲线图,求下列各式:(a)$F(3)$,(b)$f(3)$,(c)$P(X>3)$,(d)$F(3)-F(2)$,(e)$F(3.8)$,(f)$f(3.8)$.

解☞

(a)$F(3)=0.4$.

(b)由 10.7 节可以知道 3 的概率或 $f(3)$,就是在 3 这点 $F(x)$的阶梯步长,因此 $f(x)=0.2$

(c)$P(X>3)$,即 X 取值大于 3 的概率等于$\sum_{x>3}f(x)=f(4)+f(5)+f(6)$,也就是 $1-F(3)$. 因此,

$$P(X>3)=1-F(3)=1-0.4=0.6$$

(d)$F(3)-F(2)=f(3)=0.2$

(e)$F(3.8)=0.4$

(f)$f(3.8)=0$

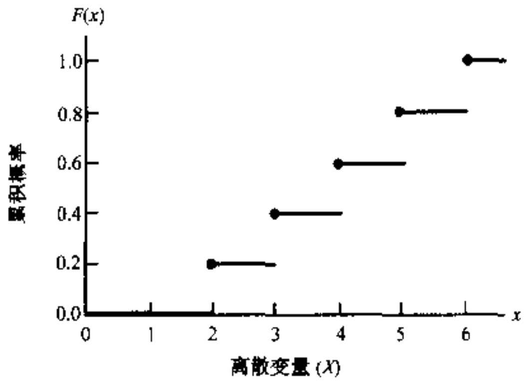


图 10-13

10.8 根据图 10-13 所示的累积分布函数,求下列各式(a)$P(3<X\leqslant5)$,(b)$P(3\leqslant X\leqslant5)$,(c)$P(3<X<5)$,(d)$P(3\leqslant X\leqslant5)$.

解☞　图 10-13 中离散型随机变量 X 可以取 5 个值:$x=2,3,4,5,6$. 这些事件是互不相容的,所以可以用集合论性质 4(见 8.6 节),求事件的并的概率.

$$P(A_1\cup A_2\cup\cdots\cup A_k)=P(A_1)+P(A_2)+\cdots P(A_k)$$

(a) $P(3<X\leqslant5)=P(4\cup5)=f(4)+f(5)=F(5)-F(3)=0.8-0.4=0.4$

(b) $P(3\leqslant X\leqslant5)=P(3\cup4\cup5)=f(3)+f(4)+f(5)=F(5)-F(3)+f(3)=0.8-0.4+0.2=0.6$

(c) $P(3<X<5)=P(4)=f(4)=F(5)-F(3)-f(5)=0.8-0.4-0.2=0.2$

(d) $P(3\leqslant X<5)=P(3\cup4)=f(3)+f(4)=F(5)-F(3)+f(3)-f(5)=0.8-0.4+0.2-0.2=0.4$

随机变量的期望值

10.9 在掷骰子试验中,设随机变量为出现的点数(见 10.3 节),求它的期望.

解☞ 利用(10.10)式对表10.1,求离散概率分布的期望

$$E(X)=\sum_x xf(x)=\left(1\times\frac{1}{6}\right)+\left(2\times\frac{1}{6}\right)+\left(3\times\frac{1}{6}\right)$$
$$+\left(4\times\frac{1}{6}\right)+\left(5\times\frac{1}{6}\right)+\left(6\times\frac{1}{6}\right)$$
$$=(0.166667)+(0.333333)+(0.5)+(0.666667)+(0.833333)+(1.0)$$
$$=3.5$$

这也就是说,如果试验重复很多次,期望平均结果是3.5点.

10.10 连掷两次骰子,设随机变量为两次所出现的总点数,其概率分布见表10.3,求其期望值.

解☞ 用表10.3求期望$E(X)$,即求所有$xf(x)$的和,表10.6给出了结果和$E(X)$的计算值,$E(X)=7.0$,表明若大量重复此试验,得到平均出现7.0点.

表 10.6

点数 x	概率 $f(x)$	$xf(x)$	点数 x	概率 $f(x)$	$xf(x)$
2	0.02778	0.05556	8	0.13890	1.11120
3	0.05556	0.16668	9	0.11112	1.00008
4	0.08334	0.33336	10	0.08334	0.83340
5	0.11112	0.55560	11	0.05556	0.61116
6	0.13890	0.83340	12	0.02778	0.33336
7	0.16668	1.16676	$\sum$	1.00	7.00056,或7.0

10.11 在8.8节指出,对给定的一次赌博,如果事件发生的机率是均等的,则这是一次公平的赌博,下面用期望来定义公平赌博.

如果X是一个离散型随机变量,它表示一个试验的可能结果,在一次赌博中损失x_1,$x_2,\cdots,x_k$,的概率分别是$f(x_1),f(x_2),\cdots,f(x_k)$.若$E(X)=\sum_x xf(x)=0$则此赌博认为是公平赌局.

实际上,这意味着如果此试验或赌博重复很多次,从长远来看结果是平局,也就是既不赢也不输,那么它就是一个公平赌博.由期望值定义,判断下列哪种情况可看作公平赌局.(a)在抛硬币时,若抛出正面你就赢\$3,若是反面就输\$2.(b)掷一次骰子,若抛出偶数点你就赢\$3,奇数点就输\$3.(c)从一副扑克牌中抽一张牌,若是红牌你赢\$1,若是黑牌输\$2.

解☞

(a) $E(X)=\sum_x f(x)=\left(\$3\times\frac{1}{2}\right)+\left(-\$2\times\frac{1}{2}\right)=\$1.5-\$1=\$0.5$

这不是公平赌博,因为从长远来看每一次投掷赢的期望是0.5美元.

(b) $$E(X)=\sum_x xf(x)=\left(-\$3\times\frac{1}{6}\right)+\left(\$3\times\frac{1}{6}\right)+\left(-\$3\times\frac{1}{6}\right)+\left(\$3\times\frac{1}{6}\right)$$
$$+\left(-\$3\times\frac{1}{6}\right)+\left(\$3\times\frac{1}{6}\right)=3\left(-\frac{\$3}{6}\right)+3\left(\frac{\$3}{6}\right)$$
$$=-\$1.5+\$1.5=\$0$$

这是公平赌博.从长远看你将不输不赢.

(c) $E(X)=\sum_x xf(x)=\left(\$1\times\frac{26}{52}\right)+\left(-\$2\times\frac{26}{52}\right)=\$0.5-\$1.0=-\$0.5$

这不是公平赌博,从长远看平均每局输\$0.5

10.12 一个慈善机构正在举办抽奖活动,奖项如下:一个奖金\$500,五个\$100,五十个\$50,他们计划卖出5000张彩票,如果彩票价格是公平价格的3倍,你建议票价是多少?

解☞ 首先,求获得每项奖金的概率$f(\$500)=1/5000=0.0002$;$f(\$100)=5/5000=0.001$;

$f(\$50)=50/5000 \quad 0.01$. 每张票的期望值为

$$E(X)=\sum_x xf(x)=(\$500\times0.0002)+(\$100\times0.001)+(\$50\times0.01)$$

$$=\$0.10+\$0.10+\$0.50=\$0.70$$

因此，每张彩票的公平价格是 \$0.70，但为了使慈善机构获得盈利应建议每张彩票售价 \$2.10.

10.13 表 10.7 是一个电子商店四年中每周售出的最受欢迎的计算机，编制清单，商店经理想知道在未来的 6 个月内商店能卖出多少台这样的计算机.

表 10.7

每周计算机的售量 x_i	周数 f_i	每周计算机的售量 x_i	周数 f_i
1	6	5	25
2	33	6	17
3	50	7	7
4	70	$\sum$	208

解 ☞ 首先设离散型变量为每周卖出的计算机数量，求它的期望值，需要列出含有 $f(x)$ 与 $xf(x)$ 列的表，$E(X)$ 计算结果. 因此在表 10.8 中给出在未来的 6 个月里商店可卖出 $26\times[E(X)=3.740385]=97.250010$ 或 97 台计算机.

表 10.8

每周计算机售量 x	概率 $f(x)$	$xf(x)$	每周计算机售量 x	概率 $f(x)$	$xf(x)$
1	0.028846	0.028846	5	0.120192	0.600960
2	0.158654	0.317308	6	0.081731	0.490386
3	0.240385	0.721155	7	0.033654	0.235578
4	0.336538	1.346152	$\sum$	3.740385	1.00

随机变量的方差和标准差

10.14 连续掷一枚硬币三次，设出现正面的次数为随机变量，用(10.22)和(10.23)式，定义出变量的方差和标准差.

解 ☞ 首先，必须求期望值，把表 10.2 中的数据代入(10.10)式中

$$E(X)=\sum_x xf(x)=(0\times0.125)+(1\times0.375)+(2\times0.375)+(3\times0.125)$$

$$=(0)+(0.375)+(0.750)+(0.375)=1.5$$

这样，$E(X)=\mu=1.5$. 然后把表 10.2 的数据代入(10.22)式

$$\sigma^2=\sum_x x^2f(x)-\mu^2$$

$$=[(0\times0.125)+(1\times0.375)+(4\times0.375)+(9\times0.125)]-(1.5)^2$$

$$=0.75$$

因此，由(10.23)式有标准差

$$\sigma=\sqrt{\sigma^2}=\sqrt{0.75}=0.87$$

10.15 连续两次掷骰子，设两次骰子所出现的点数为随机变量，其概率分布如表 10.3 所示，利用(10.22)和(10.23)式，求其方差和标准差.

解 ☞ 利用表 10.3 确定 σ^2 和 σ，就是把 x^2 和 $x^2f(x)$ 两列各项相加，在表 10.4 中给出其结果和 σ^2 和 σ 的计算值.

10.16 在计算机出售试验中，设每周出售的计算机台数为随机变量，其分布如表 10.8，利用(10.22)和(10.23)式，求方差和标准差.

解 ☞ 表 10.9 给出两列 x^2 和 $x^2f(x)$ 及 σ^2 和 σ 的计算结果.

表 10.9

每周计算机的售量 x	x^2	概率 $f(x)$	$xf(x)$	$x^2f(x)$
1	1	0.028846	0.028846	0.028846
2	4	0.158654	0.317308	0.634616
3	9	0.240385	0.721155	2.163465
4	16	0.336538	1.346152	5.384608
5	25	0.120192	0.600960	3.004800
6	36	0.081731	0.490386	2.942316
7	49	0.033654	0.235578	1.649046
$\sum$		1.00	3.740385	15.807697

$$\sigma^2=\sum x^2f(x)-\mu^2=15.807697-(3.740385)^2=15.807697-13.990480=1.817217\text{，或 }1.82$$

$$\sigma=\sqrt{\sigma^2}=\sqrt{1.817217}=1.348042\text{，或 }1.35$$

Chebyshev 定理和经验法则

10.17 在计算机出售试验中，设每周出售的计算机台数为离散型随机变量. 此随机变量在区间 $\mu\pm2\sigma$ 上取值的概率是多少?

解 ☞ 由 Chebyshev 定理(见 10.14 节)，则所求概率至少为

$$1-\frac{1}{k^2}=1-\frac{1}{2^2}=0.75$$

由例 10.12，精确计算知道 $\mu=3.74$ 和 $\sigma=1.35$(见表 10.9)

$$\mu\pm2\sigma \text{ 是 } 3.74\pm2(1.35),\quad \text{或 } 3.74\pm2.70,\quad \text{或 } 1.04 \text{ 或 } 6.44$$

由于在此区间上，X 可以从 2 到 6 之间取值，这样 X 的概率分布如表 10.8 和 10.9 所示的，则 X 取这些值中任意一个的概率为

$$\sum_{x=2}^{6}f(x)=0.937500\quad \text{或 } 0.94$$

把此离散型变量当作连续型变量处理，其分布也近似地看作正态分布，那么应用经验准则(见 10.14 节)，则有 X 在区间 $\mu\pm2\sigma$ 上取值的概率约为 0.95.

补 充 习 题

随机变量和概率分布

10.18 指出下列各随机变量是离散型的还是连续型的：(a) 每小时经过交叉路口的小汽车数量，(b)每辆汽车在停车标志牌前的停车时间，(c)一碗粥里的含糖量(单位是克)，(d) 每天日照时间，(e) 某医院一年中出生的婴儿人数，(f) 牧场里树的高度.

答案 (a)离散型随机变量，(b)连续型随机变量，(c)连续型随机变量，(d)连续型随机变量，(e)离散型随机变量，(f)连续型随机变量.

10.19 在一次公司的晚宴上，把 50 张纸币放在一顶帽子里，有 3 张 1 美元的，10 张 5 美元的，23 张 10 美元的，13 张 20 美元的，13 张 50 美元的。这些纸币被充分混合，一个职员从中任意取出一张纸币，取出不同纸币的概率各是多少?

答案 $f(\$1)=0.06$，$f(\$5)=0.20$，$f(\$10)=0.46$，$f(\$20)=0.26$，$f(\$50)=0.02$

10.20 一个社会学家想调查某 815 座房子中每一座房中所居住人数，发现有 93 座房只有一人住，160 座房中有 2 人住，320 座房中有 3 人住，居住 4 人的有 10 座，有 5 人的 82 座，有 6 人的有 50 座，若对每座房子的选取是随机的，则每座房中所住人数的概率是多少?

答案 $f(1)=0.114$，$f(2)=0.196$，$f(3)=0.393$，$f(4)=0.135$，$f(5)=0.101$，$f(6)=0.061$

10.21 一鸟类专家想知道某岛上黄莺在每个巢中的产蛋数，在 150 个黄莺巢中，发现有 4 个巢中 2 个蛋，36

个巢中有 3 个蛋，66 个巢有 4 个蛋，40 个巢有 5 个蛋，4 个巢有 6 个蛋，若对每个巢的选取是随机的，则巢中蛋数的概率如何？

答案 $f(2)=0.027$, $f(3)=0.240$, $f(4)=0.440$, $f(5)=0.267$, $f(6)=0.027$

随机变量的累积分布函数

10.22 连续三次抛一枚硬币，设出现正面的次数为随机变量，根据例 10.5(其分布函数见表 10.2)，并把它转化为累积分布函数[参看图 10-7(b)]，用简化形式表述此函数并作图.

答案 结果分别见图 10-14(a)和(b).

(a)

$$F(x)=\begin{cases}0 & x<0\\ 0.125 & 0\leqslant x<1\\ 0.500 & 1\leqslant x<2\\ 0.875 & 2\leqslant x<3\\ 1.000 & 3\leqslant x\end{cases}$$

(b)

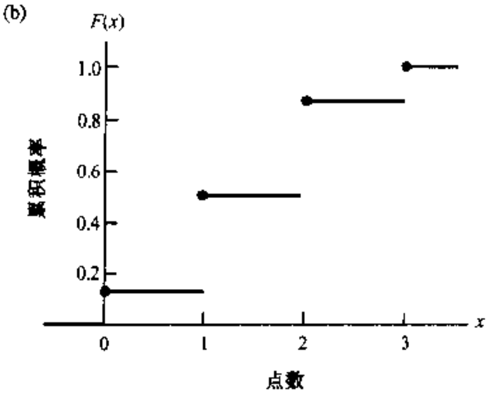


图 10-14

10.23 由图 10-15 所给的累积分布函数，求下列各式：(a)$F(4)$，(b)$f(4)$，(c)$P(X\geqslant 3)$，(d)$F(5)-F(4)$，(e)$F(4.4)$，(f)$f(4.4)$.

答案 (a)$F(4)=0.45$, (b)$f(4)=0.20$, (c)$P(X\geqslant 3)=\sum_{x\geqslant 3}f(x)=f(3)+f(4)+f(5)f(6)=1-F(2)$；因此 $P(X\geqslant 3)=1-F(2)=1-0.10=0.90$, (d)$F(5)-F(4)=f(5)=0.25$, (e)$F(4.4)=0.45$, (f)$f(4.4)=0$

$$F(x)=\begin{cases}0 & x<2\\ 0.10 & 2\leqslant x<3\\ 0.25 & 3\leqslant x<4\\ 0.45 & 4\leqslant x<5\\ 0.70 & 5\leqslant x<6\\ 1.00 & 6\leqslant x\end{cases}$$

图 10-15

10.24 见问题 10.19 从帽子中取纸币试验中，由概率[$f(x)$]写出其累积分布函数[$F(x)$]，并按从小到大的顺序排列.

答案 F($1)=0.06, F($5)=0.26, F($10)=0.72, F($20)=0.98, F($50)=1.00

10.25 见问题 10.20 房屋居住人数试验中，由概率[$f(x)$]写出其累积分布函数[$F(x)$]，并按从小到大的顺序排列.

答案 $F(1)=0.114$, $F(2)=0.310$, $F(3)=0.703$, $F(4)=0.838$, $F(5)=0.939$, $F(6)=1.000$

10.26 见问题 10.21 巢中有黄莺蛋数试验中，由概率[$f(x)$]写出其累积分布函数[$F(x)$]，并按从小到大的顺序排列.

答案 $F(2)=0.027$, $F(3)=0.267$, $F(4)=0.707$, $F(5)=0.974$, $F(6)=1.001$，或 1

随机变量的期望值

10.27 连续三次抛一枚硬币，设出现正面的次数为随机变量，若此试验重复 500 次，则希望出现多少次正面？

答案 750

10.28 下面这场赌博的公平价为多少：掷一次骰子，若出现一点或六点就得 6 美元，出现 2 点或 5 点就得 3

美元,出现了 3 点或 4 点就不得?(详细解参看问题 10.11.)

答案 $3

10.29 一个保险代理人卖给一个 35 岁妇女 10,000 美元的生活保险.要求她每年付 130 美元的保险费,从这位妇女过去记录知,她在 35 岁到 36 岁之间死亡的概率为 3/1000,那么从这份保险中公司能获利多少?

答案 $100.00

10.30 见问题 10.19 从帽子中取纸币试验中,若有放回地从帽子里取很多次,则平均起来看,此公司每次要付多少钱?

答案 $11.86

10.31 见问题 10.20 房屋居住人数试验中,若这些试验重复很多次,且每次都从 815 座房中选取.平均起来看,则在一座房子里有多少人住?

答案 3.10

10.32 见问题 10.21 巢中黄莺蛋个数试验中,若试验重复很多次,且每次都从 150 个巢中选取,则平均起来一个巢中有多少蛋?

答案 4.03

随机变量的方差和标准差

10.33 若问题 10.30 是可重复试验,求每次取钱数的方差和标准差?

答案 $\sigma^2=64.400$, $\sigma=8.02$

10.34 若问题 10.31 是可重复试验,求每座房所住人数的方差与标准差?

答案 $\sigma^2=1.706$, $\sigma=1.31$

10.35 设问题 10.32 是可重复试验,求每个巢中的蛋数的方差与标准差?

答案 $\sigma^2=0.714$, $\sigma=0.85$

Chebyshev 定理和经验法则

10.36 问题 10.31 和 10.34 是可重复试验,当用 Chebyshev 定理时,随机变量(每座房中人数)在区间 $\mu\pm2\sigma$ 取值的概率至少为 0.75.当用经验法则时,如果分布为正态分布时,则此概率约为 0.95,那么精确概率为多少?

答案 $P=0.939$

10.37 设问题 10.32 和 10.5 是可重复试验,当用 Chebyshev 定理时,随机变量(巢中产蛋数)在区间 $\mu\pm2\sigma$ 取值的概率至少是 0.75,当用经验法则时,若分布为正态分布时,此概率约为 0.95,那么精确概率为多少?

答案 $P=0.947$

附　录

表 A.1　随机数表

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	10	09	73	25	33	76	52	01	35	86	34	67	35	48	76	80	95	90	91	17	39	29	27	49	45
2	37	54	20	48	05	64	89	47	42	96	24	80	52	40	37	20	63	61	04	02	00	82	29	16	65
3	08	42	26	89	53	19	64	50	93	03	23	20	90	25	60	15	95	33	47	64	35	08	03	36	06
4	99	01	90	25	29	09	37	67	07	15	38	31	13	11	65	88	67	67	43	97	04	43	62	76	59
5	12	80	79	99	70	80	15	73	61	47	64	03	23	66	53	98	95	11	68	77	12	17	17	68	33
6	66	06	57	47	17	34	07	27	68	50	36	69	73	61	70	65	81	33	98	85	11	19	92	91	70
7	31	06	01	08	05	45	57	18	24	06	35	30	34	26	14	86	79	90	74	39	23	40	30	97	32
8	85	26	97	76	02	02	05	16	56	92	68	66	57	48	18	73	05	38	52	47	18	62	38	85	79
9	63	57	33	21	35	05	32	54	70	48	90	55	35	75	48	28	46	82	87	09	83	49	12	56	24
10	73	79	64	57	53	03	52	96	47	78	35	80	83	42	82	60	93	52	03	44	35	27	38	84	35
11	98	52	01	77	67	14	90	56	86	07	22	10	94	05	58	60	97	09	34	33	50	50	07	39	98
12	11	80	50	54	31	39	80	82	77	32	50	72	56	82	48	29	40	52	42	01	52	77	56	78	51
13	83	45	29	96	34	06	28	89	80	83	13	74	67	00	78	18	47	54	06	10	68	71	17	78	17
14	88	68	54	02	00	86	50	75	84	01	36	76	66	79	51	90	36	47	64	93	29	60	91	10	62
15	99	59	46	73	48	87	51	76	49	69	91	82	60	89	28	93	78	56	13	68	23	47	83	41	13
16	65	48	11	76	74	17	46	85	09	50	58	04	77	69	74	73	03	95	71	86	40	21	81	65	44
17	80	12	43	56	35	17	72	70	80	15	45	31	82	23	74	21	11	57	82	53	14	38	55	37	63
18	74	35	09	98	17	77	40	27	72	14	43	23	60	02	10	45	52	16	42	37	96	28	60	26	55
19	69	91	62	68	03	66	25	22	91	48	36	93	68	72	03	76	62	11	39	90	94	40	05	64	18
20	09	89	32	05	05	14	22	56	85	14	46	42	75	67	88	96	29	77	88	22	54	38	21	45	98
21	91	49	91	45	23	68	47	92	76	86	46	16	28	35	54	94	75	08	99	23	37	08	92	00	48
22	80	33	69	45	98	26	94	03	68	58	70	29	73	41	35	53	14	03	33	40	42	05	08	23	41
23	44	10	48	19	49	85	15	74	79	54	32	97	92	65	75	57	60	04	08	81	22	22	20	64	13
24	12	55	07	37	42	11	10	00	20	40	12	86	07	46	97	96	64	48	94	39	28	70	72	58	15
25	63	60	64	93	29	16	50	53	44	84	40	21	95	25	63	43	65	17	70	82	07	20	73	17	90
26	61	19	69	04	46	26	45	74	77	74	51	92	43	37	29	65	39	45	95	93	42	58	26	05	27
27	15	47	44	52	66	95	27	07	99	53	59	36	78	38	48	82	39	61	01	18	33	21	15	94	66
28	94	55	72	85	73	67	89	75	43	87	54	62	24	44	31	91	19	04	25	92	92	92	74	59	73
29	42	48	11	62	13	97	34	40	87	21	16	86	84	87	67	03	07	11	20	59	25	70	14	66	70
30	23	52	37	83	17	73	20	88	98	37	68	93	59	14	16	26	25	22	96	63	05	52	28	25	62
31	04	49	35	24	94	75	24	63	38	24	45	86	25	10	25	61	96	27	93	35	65	33	71	24	72
32	00	54	99	76	54	64	05	18	81	59	96	11	96	38	96	54	69	28	23	91	23	28	72	95	29
33	35	96	31	53	07	26	89	80	93	54	33	35	13	54	62	77	97	45	00	24	90	10	33	93	33
34	59	80	80	83	91	45	42	72	68	42	83	60	94	97	00	13	02	12	48	92	78	56	52	01	06
35	46	05	88	52	36	01	39	09	22	86	77	28	14	40	77	93	91	08	36	47	70	61	74	29	41
36	32	17	90	05	97	87	37	92	52	41	05	56	70	70	07	86	74	31	71	57	85	39	41	18	38
37	69	23	46	14	06	20	11	74	52	04	15	95	66	00	00	18	74	39	24	23	97	11	89	63	38
38	19	56	54	14	30	01	75	87	53	79	40	41	92	15	85	66	67	43	68	06	84	96	28	52	07
39	45	15	51	49	38	19	47	60	72	46	43	66	79	45	43	59	04	79	00	33	20	82	66	95	41
40	94	86	43	19	94	36	16	81	08	51	34	88	88	15	53	01	54	03	54	56	05	01	45	11	76

续表

	26	27	28	29	30	3	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	98	08	62	48	26	45	24	02	84	04	44	99	90	88	96	39	09	47	34	07	35	44	13	18	80
2	33	18	51	62	32	41	94	15	09	49	89	43	54	85	81	88	69	54	19	94	37	54	87	30	43
3	80	95	10	04	06	96	38	27	07	74	20	15	12	33	87	25	01	62	52	98	94	62	46	11	71
4	79	75	24	91	40	71	96	12	82	96	69	86	10	25	91	74	85	22	05	39	00	38	75	95	79
5	18	63	33	25	37	98	14	50	65	71	31	01	02	46	74	05	45	56	14	27	77	93	89	19	36
6	74	02	94	39	02	77	55	73	22	70	97	79	01	71	19	52	52	75	80	21	80	81	45	17	48
7	54	17	84	56	11	80	99	33	71	43	05	33	51	29	69	56	12	71	92	55	36	04	09	03	24
8	11	66	44	98	83	52	07	98	48	27	59	38	17	15	39	09	97	33	34	40	88	46	12	33	56
9	48	32	47	79	28	31	24	96	47	10	02	29	53	68	70	32	30	75	75	46	15	02	00	99	94
10	69	07	49	41	38	87	63	79	19	76	35	58	40	44	01	10	51	82	16	15	01	84	87	69	38
11	09	18	82	00	97	32	82	53	95	27	04	22	08	63	04	83	38	98	73	74	64	27	85	80	44
12	90	04	58	54	97	51	98	15	06	54	94	93	88	19	97	91	87	07	61	50	68	47	66	46	59
13	73	18	95	02	07	47	67	72	52	69	62	29	06	44	64	27	12	46	70	18	41	36	18	27	60
14	75	76	87	64	90	20	97	18	17	49	90	42	91	22	72	95	37	50	58	71	93	82	34	31	78
15	54	01	64	40	56	66	28	13	10	03	00	68	22	73	98	20	71	45	32	95	07	70	61	78	13
16	08	35	86	99	10	78	54	24	27	85	13	66	15	88	73	04	61	89	75	53	31	22	30	84	20
17	28	30	60	32	64	81	33	31	05	91	40	51	00	78	93	32	60	46	04	75	94	11	90	18	40
18	53	84	08	62	33	81	59	41	36	28	51	21	59	02	90	28	46	66	87	95	77	76	22	07	91
19	91	75	75	37	41	61	61	36	22	69	50	26	39	02	12	55	78	17	65	14	83	48	34	70	55
20	89	41	59	26	94	00	39	75	83	91	12	60	71	76	46	48	94	97	23	06	94	54	13	74	08
21	77	51	30	38	20	86	83	42	99	01	68	41	48	27	74	51	90	81	39	80	72	89	35	55	07
22	19	50	23	71	74	69	97	92	02	88	55	21	02	97	73	74	28	77	52	51	65	34	46	74	15
23	21	81	85	93	13	93	27	88	17	57	05	68	67	31	56	07	08	28	50	46	31	85	33	84	52
24	51	47	46	64	99	68	10	72	36	21	94	04	99	13	45	42	83	60	91	91	08	00	74	54	49
25	99	55	96	83	31	62	53	52	41	70	69	77	71	28	30	74	81	97	81	42	43	86	07	28	34
26	33	71	34	80	07	93	58	47	28	69	51	92	66	47	21	58	30	32	98	22	93	17	49	39	72
27	85	27	48	68	93	11	30	32	92	70	28	83	43	41	37	73	51	59	04	00	71	14	84	36	43
28	84	13	38	96	40	44	03	55	21	66	73	85	27	00	91	61	22	26	05	61	62	32	71	84	23
29	56	73	21	62	34	17	39	59	61	31	10	12	39	16	22	85	49	65	75	60	81	60	41	88	80
30	65	13	85	68	06	87	64	88	52	61	34	31	36	58	61	45	87	52	10	69	85	64	44	72	77
31	38	00	10	21	76	81	71	91	17	11	71	60	29	29	37	74	21	96	40	49	65	58	44	96	98
32	37	40	29	63	97	01	30	47	75	86	56	27	11	00	86	47	32	46	26	05	40	03	03	74	38
33	97	12	54	03	48	87	08	33	14	17	21	81	53	92	50	75	23	76	20	47	15	50	12	95	78
34	21	82	64	11	34	47	14	33	40	72	64	63	88	59	02	49	13	90	64	41	03	85	65	45	52
35	73	13	54	27	42	95	71	90	90	35	85	79	47	42	96	08	78	98	81	56	64	69	11	92	02
36	07	63	87	79	29	03	06	11	80	72	96	20	74	41	56	23	82	19	95	38	04	71	36	69	94
37	60	52	88	34	41	07	95	41	98	14	59	17	52	06	95	05	53	35	21	39	61	21	20	64	55
38	83	59	63	56	55	06	95	89	29	83	05	12	80	97	19	77	43	35	37	83	92	30	15	04	98
39	10	85	06	27	46	99	59	91	05	07	13	49	90	63	19	53	07	57	18	39	06	41	01	93	62
40	39	82	09	89	52	43	62	26	31	47	64	42	18	08	14	43	80	00	93	51	31	02	47	31	67

续表

	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
1	59	58	00	64	78	75	56	97	88	00	88	83	55	44	86	23	76	80	61	56	04	11	10	84	08
2	38	50	80	73	41	23	79	34	87	63	90	82	29	70	22	17	71	90	42	07	95	95	44	99	53
3	30	69	27	06	68	94	68	81	61	27	56	19	68	00	91	82	06	76	34	00	05	46	26	92	00
4	65	44	39	56	59	18	28	82	74	37	49	63	22	40	41	08	33	76	56	76	96	29	99	08	36
5	27	26	75	02	64	13	19	27	22	94	07	47	74	46	06	17	98	54	89	11	97	34	13	03	58
6	91	30	70	69	91	19	07	22	42	10	36	69	95	37	28	28	82	53	57	93	28	97	66	62	52
7	68	43	49	46	88	84	47	31	36	22	62	12	69	84	08	12	84	38	25	90	09	81	59	31	46
8	48	90	81	58	77	54	74	52	45	91	35	70	00	47	54	83	82	45	26	92	54	13	05	51	60
9	06	91	34	51	97	42	67	27	86	01	11	88	30	95	28	63	01	19	89	01	14	97	44	03	44
10	10	45	51	60	19	14	21	03	37	12	91	34	23	78	21	88	32	58	08	51	43	66	77	08	83
11	12	88	39	73	43	65	02	76	11	84	04	28	50	13	92	17	97	41	50	77	90	71	22	67	69
12	21	77	83	09	76	38	80	73	69	61	31	64	94	20	96	63	28	10	20	23	08	81	64	74	49
13	19	52	35	95	15	65	12	25	96	59	86	28	36	82	58	69	57	21	37	98	16	43	59	15	29
14	67	24	55	26	70	35	58	31	65	63	79	24	68	66	86	76	46	33	42	22	26	65	59	08	02
15	60	58	44	73	77	07	50	03	79	92	45	13	42	65	29	26	76	08	36	37	41	32	64	43	44
16	53	85	34	13	77	36	06	69	48	50	58	83	87	38	59	49	36	47	33	31	96	24	04	36	42
17	24	63	73	87	36	74	38	48	93	42	52	62	30	79	92	12	36	91	86	01	03	74	28	38	73
18	83	08	01	24	51	38	99	22	28	15	07	75	95	17	77	97	37	72	75	85	51	97	23	78	67
19	16	44	42	43	34	36	15	19	90	73	27	49	37	09	39	85	13	03	25	52	54	84	65	47	59
20	60	79	01	81	57	57	17	86	57	62	11	16	17	85	76	45	81	95	29	79	65	13	00	48	60
21	03	99	11	04	61	93	71	61	68	94	66	08	32	46	53	84	60	95	82	32	88	61	81	91	61
22	38	55	59	55	54	32	88	65	97	80	08	35	56	08	60	29	73	54	77	62	71	29	92	38	53
23	17	54	67	37	04	92	05	24	62	15	55	12	12	92	81	59	07	60	79	36	27	95	45	89	09
24	32	64	35	28	61	95	81	90	68	31	00	91	19	89	36	76	35	59	37	79	80	86	30	05	14
25	69	57	26	87	77	39	51	03	59	05	14	06	04	06	19	29	54	96	96	16	33	56	46	07	80
26	24	12	26	65	91	27	69	90	64	94	14	84	54	66	72	61	95	87	71	00	90	89	97	57	54
27	61	19	63	02	31	92	96	26	17	73	41	83	95	53	82	17	26	77	09	43	78	03	87	02	67
28	30	53	22	17	04	10	27	41	22	02	39	68	52	33	09	10	06	16	88	29	55	98	66	64	85
29	03	78	89	75	99	75	86	72	07	17	74	41	65	31	66	35	20	83	33	74	87	53	90	88	23
30	48	22	86	33	79	85	78	34	76	19	53	15	26	74	33	35	66	35	29	72	16	81	86	03	11
31	60	36	59	46	53	35	07	53	39	49	42	61	42	92	97	01	91	82	83	16	98	95	37	52	31
32	83	79	94	24	02	56	62	33	44	42	34	99	44	13	74	70	07	11	47	36	09	95	81	80	65
33	32	96	00	74	05	36	40	98	32	32	99	38	54	16	00	11	13	30	75	86	15	91	70	62	53
34	19	32	25	38	45	57	62	05	26	06	66	49	76	86	46	78	13	86	65	59	19	64	09	94	13
35	11	22	09	47	47	07	39	93	74	08	48	50	92	39	29	27	48	24	54	76	85	24	43	51	59
36	31	75	15	72	60	68	98	00	53	39	15	47	04	83	55	88	65	12	25	96	03	15	21	92	21
37	88	49	29	93	82	14	45	40	45	04	20	09	49	89	77	74	84	39	34	13	22	10	97	85	08
38	30	93	44	77	44	07	48	18	38	28	73	78	80	65	33	28	59	72	04	05	94	20	52	03	80
39	22	88	84	88	93	27	49	99	87	48	60	53	04	51	28	74	02	28	46	17	82	03	71	02	68
40	78	21	21	69	93	35	90	29	13	86	44	37	21	54	86	65	74	11	40	14	87	48	13	72	20

表 A.2　统计分类数据

本表给出例 3.5 中介绍的统计专业数据.分二部分(女生和男生)给出:学生姓名的第一个字母(第 1 列);学号(第 2 列);期末考试的百分制成绩(第 3 列);精确到 1/4 英寸的身高(第 4 列);体重(磅)(第 5 列);家庭收入(精确到 100 美元)(第 6 列);头发颜色(黑,金黄,棕,红)(第 7 列);学期论文等级(A,B,C,D,F)(第 8 列);是否选入(*)简单随机样本(SRS)(第 9 列);比例分层随机样本(PSRS)(第 10 列);系统随机样本(SYRS)(第 11 列);或单阶段整群样本(SCRS)(第 12 列).

姓名 (1)	学号 (2)	成绩 (3)	身高 (4)	体重 (5)	家庭收入 (6)	头发颜色 (7)	论文 (8)	SRS (9)	PSRS (10)	SYRS (11)	SCRS (12)
						女					
LB	03	83	67.75	127	31,500	棕	B	—	—	*	
AA	09	88	60.25	109	25,600	金黄	B	—	—	—	*
AE	17	57	63.75	117	76,500	黑	F	*	*	—	
MJ	27	78	65.25	123	20,200	金黄	B	*	*	*	*
NO	28	97	62.00	105	37,800	金黄	A	·	—	—	*
LT	30	82	63.50	119	15,400	黑	B	—		—	*
JD	33	91	65.25	129	71,800	金黄	A	—	—	—	—
DD	34	90	65.50	123	31,700	金黄	B	—	—		—
EF	44	80	65.25	124	30,100	棕	B	*	—	—	—
AC	49	64	64.75	121	34,700	棕	C	—	*	—	—
JH	51	87	67.00	134	40,500	金黄	A	—	*	*	—
AH	53	79	64.25	115	36,900	黑	C	*	—	—	—
GY	56	91	69.25	136	20,400	红	A	—	—	—	—
MZ	57	79	66.25	131	30,400	棕	C	—	—	—	—
BJ	58	65	63,00	111	28,500	金黄	D	—	—	—	—
TM	64	94	64.75	121	46,100	金黄	A	*	—	—	—
						男					
CA	01	90	69.25	180	21,200	棕	B	—	*	—	*
FE	02	94	65.25	138	145,000	棕	A	—	—	—	*
HE	04	59	69.00	152	29,300	棕	D	—	—	—	—
LW	05	91	73.00	172	26,600	金黄	B	—	—	—	—
OA	06	84	69.25	163	20,900	金黄	B	—	*	—	—
PS	07	96	70.25	170	26,200	金黄	A	*	*	*	—
OF	08	84	67.00	158	33,700	金黄	B	—	—	—	—
HC	10	79	71.75	190	54,200	棕	C	—	—	—	*
EB	11	84	66.25	148	28,600	棕	B	—	—	*	—
MA	12	90	66.25	157	29,200	黑	B	—	—	—	—
ME	13	72	70.25	156	58,400	棕	C	*	—	—	—
HK	14	93	68.00	164	21,700	黑	A	—	—	—	—
AD	15	69	72.00	175	27,700	黑	C	·		*	—
RE	16	87	69.00	160	24,200	金黄	B	-			-
FA	18	93	72.50	172	42,500	金黄	A	—	—	—	—
CE	19	90	65.25	152	28,100	棕	B	—	—	*	*
BP	20	74	72.25	184	22,200	红	C	—	—	—	*
EO	21	88	67.25	142	24,100	金黄	B	—	—	—	*
RA	22	86	67.00	124	49,000	棕	B	*	··	—	*
DA	23	71	68.25	152	25,100	棕	C	—	—	*	—
GK	24	80	71.00	169	31,700	黑	B	—	—	—	—
JA	25	94	68.00	150	39,200	金黄	A	—	—	—	
GB	26	98	69.25	147	35,600	金黄	A	*	—	—	··
JW	29	81	67.00	140	15,700	棕	B	—	*	—	*

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	D	—	—	*	—
WA	32	86	67.50	138	14,100	棕	B	*	—	—	
NA	35	90	69.00	159	33,300	红	B	—	—	*	
SM	36	67	67.50	160	28,300	金黄	C	—	—	—	
MQ	37	85	66.25	131	30,700	棕	B	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	D	*	—	—	*
TS	39	69	71.25	186	32,400	棕	C	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	C	—		—	—
GM	41	68	64.25	143	103,600	棕	C	—	*	—	—
BC	42	99	70.75	173	17,300	红	A	—	*	—	
CI	43	95	68.25	151	37,200	棕	A	—	*	*	
JL	45	64	70.00	160	43,700	棕	D	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	A	—	*	—	*
FV	47	87	69.50	144	31,600	黑	B	—	*	*	—
DW	48	78	70.25	158	24,300	棕	B	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	F	—	—	—	—
BF	52	88	74.00	194	23,900	红	B	—	*	—	—
NP	54	76	70.75	176	49,000	黑	C	—	—	—	—
GT	55	90	70.00	166	35,000	棕	A	—	*	*	—
LR	59	49	69.00	165	58,600	棕	F	*	—	*	—
CR	60	80	68.75	161	52,100	黑	B	—	—	—	—
PB	61	91	69.00	170	25,400	黑	A		—	—	—
FJ	62	90	70.25	151	36,100	棕	A	—	—	—	—
AT	63	84	65.25	124	42,300	黑	B	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名 (1)	学号 (2)	成绩 (3)	身高 (4)	体重 (5)	家庭收入 (6)	头发颜色 (7)	论文 (8)	SRS (9)	PSRS (10)	SYRS (11)	SCRS (12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	--	*	·
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	- -	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	--	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	- -	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	- -	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	- -	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名 (1)	学号 (2)	成绩 (3)	身高 (4)	体重 (5)	家庭收入 (6)	头发颜色 (7)	论文 (8)	SRS (9)	PSRS (10)	SYRS (11)	SCRS (12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名 (1)	学号 (2)	成绩 (3)	身高 (4)	体重 (5)	家庭收入 (6)	头发颜色 (7)	论文 (8)	SRS (9)	PSRS (10)	SYRS (11)	SCRS (12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	- -	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	- -	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	- -	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	- -	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	--	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	- -	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	--	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	- -	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	- -	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	- -	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	- -	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	- -	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	- -	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	- -	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名 (1)	学号 (2)	成绩 (3)	身高 (4)	体重 (5)	家庭收入 (6)	头发颜色 (7)	论文 (8)	SRS (9)	PSRS (10)	SYRS (11)	SCRS (12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	- -	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	- -	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	- -	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	- -	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名 (1)	学号 (2)	成绩 (3)	身高 (4)	体重 (5)	家庭收入 (6)	头发颜色 (7)	论文 (8)	SRS (9)	PSRS (10)	SYRS (11)	SCRS (12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	--	*	·
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	- -	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	--	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	- -	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	- -	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	- -	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30.700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	D	—	—	*	—
WA	32	86	67.50	138	14,100	棕	B	*	—	—	
NA	35	90	69.00	159	33,300	红	B	—	—	*	
SM	36	67	67.50	160	28,300	金黄	C	—	—	—	
MQ	37	85	66.25	131	30,700	棕	B	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	D	*	—	—	*
TS	39	69	71.25	186	32,400	棕	C	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	C	—		—	—
GM	41	68	64.25	143	103,600	棕	C	—	*	—	—
BC	42	99	70.75	173	17,300	红	A	—	*	—	
CI	43	95	68.25	151	37,200	棕	A	—	*	*	
JL	45	64	70.00	160	43,700	棕	D	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	A	—	*	—	*
FV	47	87	69.50	144	31,600	黑	B	—	*	*	—
DW	48	78	70.25	158	24,300	棕	B	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	F	—	—	—	—
BF	52	88	74.00	194	23,900	红	B	—	*	—	—
NP	54	76	70.75	176	49,000	黑	C	—	—	—	—
GT	55	90	70.00	166	35,000	棕	A	—	*	*	—
LR	59	49	69.00	165	58,600	棕	F	*	—	*	—
CR	60	80	68.75	161	52,100	黑	B	—	—	—	—
PB	61	91	69.00	170	25,400	黑	A		—	—	—
FJ	62	90	70.25	151	36,100	棕	A	—	—	—	—
AT	63	84	65.25	124	42,300	黑	B	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	·
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	-

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	—
WA	32	86	67.50	138	14,100	棕	*B*	*	—	—	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	—	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	—
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	—
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—

续表

姓名	学号	成绩	身高	体重	家庭收入	头发颜色	论文	SRS	PSRS	SYRS	SCRS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HO	31	59	73.00	191	66,900	金黄	*D*	—	—	*	-
WA	32	86	67.50	138	14,100	棕	*B*	*	—	-	
NA	35	90	69.00	159	33,300	红	*B*	—	—	*	
SM	36	67	67.50	160	28,300	金黄	*C*	—	—	—	
MQ	37	85	66.25	131	30,700	棕	*B*	*	—	—	*
JT	38	64	68.50	147	25,600	金黄	*D*	*	—	—	*
TS	39	69	71.25	186	32,400	棕	*C*	*	—	*	—
MU	40	83	69.00	152	12,700	金黄	*C*	—		—	—
GM	41	68	64.25	143	103,600	棕	*C*	—	*	—	—
BC	42	99	70.75	173	17,300	红	*A*	—	*	—	
CI	43	95	68.25	151	37,200	棕	*A*	—	*	*	
JL	45	64	70.00	160	43,700	棕	*D*	*	—	—	*
JQ	46	92	68.75	149	88,000	黑	*A*	-	*	—	*
FV	47	87	69.50	144	31,600	黑	*B*	—	*	*	—
DW	48	78	70.25	158	24,300	棕	*B*	—	*	—	—
DM	50	55	64.75	119	18,000	金黄	*F*	—	—	—	—
BF	52	88	74.00	194	23,900	红	*B*	—	*	—	—
NP	54	76	70.75	176	49,000	黑	*C*	—	—	—	-
GT	55	90	70.00	166	35,000	棕	*A*	—	*	*	-
LR	59	49	69.00	165	58,600	棕	*F*	*	—	*	—
CR	60	80	68.75	161	52,100	黑	*B*	—	—	—	—
PB	61	91	69.00	170	25,400	黑	*A*		—	—	—
FJ	62	90	70.25	151	36,100	棕	*A*	—	—	—	—
AT	63	84	65.25	124	42,300	黑	*B*	*	—	*	—